

ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AD1188  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-378 <GLA>  
A:Cross-references: GB:NC 003210; PIDN:CAC98986.1; PID:gl6410311; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo0908

Query Match 5.0%; Score 7; DB 2; Length 378;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 EMNKTIS 102  
|||  
DB 214 EMNKTIS 220  
|||

RESULT 22  
AD1546  
membrane protein homolog lin0908 [imported] - *Listeria innocua* (strain Clp11262)  
C:Species: *Listeria innocua*  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AD1546  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fahi, H.; Jones, L.M.; Karst, U.; Science 294, 849-852, 2001  
A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma, O.K.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AD1546  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-378 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CAC96140.1; PID:gl6413357; GSPDB:GN00178  
A:Experimental source: strain Clp11262  
C:Genetics:  
A:Gene: lin0908

Query Match 5.0%; Score 7; DB 2; Length 378;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 EMNKTIS 102  
|||  
DB 214 EMNKTIS 220  
|||

RESULT 23  
A70647  
Probable PPE protein - *Mycobacterium tuberculosis* (strain H37RV)  
C:Species: *Mycobacterium tuberculosis*  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: A70647  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: A70647  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-409 <COL>  
A:Cross-references: GB:Z83867; GB:AL123456; NID:G3261695; PIDN:CAB06293.1; PID:el299813  
A:Experimental source: strain H37RV  
C:Genetics:

A:Gene: PPE

Query Match 5.0%; Score 7; DB 2; Length 409;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 GASAIAA 63  
|||  
DB 157 GASAIAA 163  
|||

RESULT 24  
A70932  
Probable PPE protein - *Mycobacterium tuberculosis* (strain H37RV)  
C:Species: *Mycobacterium tuberculosis*  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: A70932  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: A70932  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-409 <COL>  
A:Cross-references: GB:AL020201; GB:AL123456; NID:G3250699; PIDN:CAA17729.1; PID:el254f  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: PPE

Query Match 5.0%; Score 7; DB 2; Length 409;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 VLPCTGA 58  
|||  
DB 342 VLPCTGA 348  
|||

RESULT 25  
ZZRCL  
nodulation protein nodC - *Rhizobium leguminosarum* plasmid pRL1J1  
C:Species: *Rhizobium leguminosarum*  
C:Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 21-Jul-2000  
C:Accession: A03486  
R:Roosen, L.; Johnston, A.W.B.; Downie, J.A.  
Nucleic Acids Res. 12, 9497-9508, 1984  
A:Title: DNA sequence of the *Rhizobium leguminosarum* nodulation genes nodAB and C required for nodulation  
A:Reference number: A03482; MUID:85087952; PMID:6514582  
A:Accession: A03486  
A:Molecule type: DNA  
A:Residues: 1-424 <ROS>  
A:Cross-references: GB:X01650; NID:G46212; PIDN:CAA68619.1; PID:G46215  
C:Comment: This is one of the proteins, coded by nodulation genes, that are required for nodulation  
C:Genetics:  
A:Gene: nodC  
A:Superfamily: nodulation protein nodC  
C:Keywords: nodulation

Query Match 5.0%; Score 7; DB 1; Length 424;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 SAIATV 65  
|||  
DB 259 SAIATV 265  
|||

RESULT 26

F87108  
A:Molecule type: DNA  
A:Residues: 1-491 <STO>  
A:Cross-references: GB:AE004637; GB:AE004091; NID:g9948105; PIDN:AA05485.1; GSPDB:GN001  
C:Species: Mycobacterium leprae  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 30-Sep-2001  
C:Accession: F87108  
R: Cole, S.T.; Eigleier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Holt,  
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,  
M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Squares, R.;  
A:Title: Massive gene decay in the leprosy bacillus.  
A:Reference number: A86909; MUID:21128732; PMID:11234002  
A:Accession: F87108  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-468 <STO>  
A:Cross-references: GB:AL450380; NID:gl3093388; PIDN:CAC30547.1; GSPDB:GN00147  
C:Genetics: MUI596  
C:Superfamily: indoleacetamide hydrolase

Query Match 5.0%; Score 7; DB 2; Length 468;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 SPELREL 90  
DB 269 SPELREL 275

RESULT 27  
T28052  
hypothetical protein ZK858.7 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T28052  
R:White, S.  
submitted to the EMBL Data Library, September 1996  
A:Reference number: Z20462  
A:Accession: T28052  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-473 <NTL>  
A:Cross-references: EMBL:Z79759; PIDN:CAB02135.1; GSPDB:GN00019; CESP:ZK858.7  
C:Genetics: MUI596  
A:Map position: 1  
A:Introns: 23/3; 51/2; 177/3; 272/3; 338/3; 404/3; 456/3

Query Match 5.0%; Score 7; DB 2; Length 473;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 RLAMLRA 43  
DB 219 RLAMLRA 225

RESULT 28  
F83383  
probable flavin-binding monooxygenase PA2097 [imported] - Pseudomonas aeruginosa (strain  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: F83383  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
L.; Lory, S.; Olson, M.V.  
Nature 406, 958-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: F83383  
A:Status: preliminary

F87108  
A:Molecule type: DNA  
A:Residues: 1-491 <STO>  
A:Cross-references: GB:AE004637; GB:AE004091; NID:g9948105; PIDN:AA05485.1; GSPDB:GN001  
C:Species: Mycobacterium leprae  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 30-Sep-2001  
C:Accession: F87108  
R: Cole, S.T.; Eigleier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Holt,  
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,  
eam, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S  
A:Title: Massive gene decay in the leprosy bacillus.  
A:Reference number: A86909; MUID:21128732; PMID:11234002  
A:Accession: F87108  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-468 <STO>  
A:Cross-references: GB:AL450380; NID:gl3093388; PIDN:CAC30547.1; GSPDB:GN00147  
C:Genetics:  
A:Gene: Mui596  
C:Superfamily: indoleacetamide hydrolase

Query Match 5.0%; Score 7; DB 2; Length 468;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 SPELREL 90  
DB 269 SPELREL 275

RESULT 27  
T28052  
hypothetical protein ZK858.7 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T28052  
R:White, S.  
submitted to the EMBL Data Library, September 1996  
A:Reference number: Z20462  
A:Accession: T28052  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-473 <NTL>  
A:Cross-references: EMBL:Z79759; PIDN:CAB02135.1; GSPDB:GN00019; CESP:ZK858.7  
C:Genetics:  
A:Gene: CESP:ZK858.7  
A:Map position: 1  
A:Introns: 23/3; 51/2; 177/3; 272/3; 338/3; 404/3; 456/3

Query Match 5.0%; Score 7; DB 2; Length 473;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 RLAMLA 43  
DB 219 RLAMLA 225

RESULT 28  
F83383  
probable flavin-binding monooxygenase PA2097 [imported] - Pseudomonas aeruginosa (strain  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: F83383  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
Lory, S.; Olson, M.V.  
Nature 406, 958-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: F83383  
A:Status: preliminary

```

Db      194 GTGASAI 200
|||||
RESULT 31
A:Accession: A83453
A:Title: Probable flavin-containing monooxygenase PA1538 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: A83453
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, N.; Lory, S.; Olson, M.V.
Nature 406, 939-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: A82950; MUID:2043737; PMID:10984043
A:Accession: A83453
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-527 <STO>
A:Cross-references: GB:AE004582; GB:AE004091; NID:9947492; PIDN:AG04927.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1538

Query Match      5.0%; Score 7; DB 2; Length 527;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      55 GTGASAI 61
|||||
Db      198 GTGASAI 204
|||||

RESULT 32
A:Accession: T40151
A:Title: histidine-tRNA ligase precursor, mitochondrial - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 19-Jul-2002
C:Accession: T40151
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Devlin, K.; Churcher, C.M.
Submitted to the EMBL Data Library, March 1998
A:Reference number: Z21842
A:Accession: T40151
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-538 <WOO>
A:Cross-references: EMBL:AL022103; PIDN:CAAI7892.1; GSPDB:GN00067; SPDB:SPBC2G2.12
A:Experimental source: strain 972h-; cosmid c2G2
C:Genetics:
A:Gene: SPDB:SPBC2G2.12
A:Map position: 2
A:Genome: nuclear
C:Superfamily: human histidine-tRNA ligase; amino acid-tRNA ligase repeat homology; histidine-tRNA ligase precursor, mitochondrial
F:15-65/Domain: amino acid-tRNA ligase repeat homology <ATL>
F:66-526/Domain: histidine-tRNA ligase homology <HTL>

Query Match      5.0%; Score 7; DB 1; Length 538;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      130 RSLKSQ 136
|||||
Db      20 RSLKSQ 26
|||||

RESULT 33
A:Accession: S46458
A:Title: transcription factor tbx2 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S46458

```

---

```

R; Bollag, R.J.; Siegfried, Z.; Cebra-Thomas, J.A.; Garvey, N.; Davison, E.M.; Silver, I.
Nature Genet. 7, 383-389, 1994
A:Title: An ancient family of embryonically expressed mouse genes sharing a conserved function
A:Reference number: S46458; MUID:95004605; PMID:7920656
A:Accession: S46458
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-701 <BOL>
A:Cross-references: GB:U15566; NID:9558875; PIDN:AAC52697.1; PID:9558876
C:Genetics:
A:Gene: Tbx2
C:Superfamily: mouse transcription factor tbx2; T-box homology
F:104-285/Domain: T-box homology <TBX>

Query Match      5.0%; Score 7; DB 1; Length 701;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      128 HLRLSKS 134
|||||
Db      78 HLRLSKS 84
|||||

RESULT 34
A:Accession: G01840
A:Title: T-box protein 2 - human
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 23-Jul-1999
C:Accession: G01840
R; Campbell, C.E.
Submitted to the EMBL Data Library, May 1995
A:Reference number: G08602
A:Accession: G01840
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-702 <CAM>
A:Cross-references: EMBL:U28049; NID:9924927; PIDN:AAA73861.1; PID:9924928
C:Genetics:
A:Gene: GDB:TBX2
A:Cross-references: GDB:568496; OMIM:600747
A:Map position: 17q21-17q22
C:Superfamily: mouse transcription factor tbx2; T-box homology
F:104-285/Domain: T-box homology <TBX>

Query Match      5.0%; Score 7; DB 2; Length 702;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      128 HLRLSKS 134
|||||
Db      78 HLRLSKS 84
|||||

RESULT 35
A:Accession: AH2546
A:Title: hypothetical protein all7659 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC71201
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AH2546
R; Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, N.; Nakazaki, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, N.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AH2546
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-731 <KUR>
A:Cross-references: GB:AP003602; PIDN:BAF77302.1; PID:gl7134744; GSPDB:GN00181
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all7659

```

```

A:Genome: plasmid
Query Match          5.0%; Score 7; DB 2; Length 731;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 NKTISOE 104
    |||||
Db 434 NKTISOE 440

RESULT 36
PTS system transcription regulator enzyme I [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: F91089
R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
  gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
  DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic
  A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: F91089
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-748 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA037109.1; PID:gl3363158; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIND 050952
C:Genetics:
A:Gene: EC3686

Query Match          5.0%; Score 7; DB 2; Length 748;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 AMLRALA 45
    |||||
Db 646 AMLRALA 652

RESULT 37
PTS system transcription regulator enzyme I [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 27-Nov-2001
C:Accession: H85934
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
  Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,
  Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: H85934
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-748 <STO>
A:Cross-references: GB:AE005174; NID:gl2517314; PIDN:AAG57940.1; GSPDB:GN00145; UWGP:Z41
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: ptsP

Query Match          5.0%; Score 7; DB 2; Length 748;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 AMLRALA 45
    |||||
Db 646 AMLRALA 652

RESULT 38
phosphotransferase system enzyme I (EC 2.7.3.9) - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: F65065
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
  A.; Rose, D.J.; Mau, B.; Shao, Y.
  Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: F65065
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-748 <BLAT>
A:Cross-references: GB:AE000366; GB:U00096; NID:gl789185; PIDN:AAC75868.1; PID:gl78919
A:Experimental source: strain K-12, substrain MGL1655
C:Genetics:
A:Gene: ptsP
C:Superfamily: phosphotransferase system enzyme I homology
C:Keywords: phosphotransferase
F:171-727/Domain: phosphotransferase system enzyme I homology <PT1>

Query Match          5.0%; Score 7; DB 2; Length 748;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 AMLRALA 45
    |||||
Db 646 AMLRALA 652

RESULT 39
T32844
hypothetical protein K05F6.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32844
R:Du, Z.; Goela, D.
  submitted to the EMBL Data Library, December 1997
A:Description: The sequence of C. elegans cosmid K05F6.
A:Reference number: Z21233
A:Accession: T32844
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-753 <DUZ>
A:Cross-references: EMBL:AF040653; PIDN:AAB95025.1; GSPDB:GN00020; CESP:K05F6.5
A:Experimental source: strain Bristol N2; clone K05F6
C:Genetics:
A:Gene: CESP:K05F6.5
A:Map position: 2
A:Introns: 28/2; 297/2; 343/3; 409/2; 551/2; 596/3; 666/3

Query Match          5.0%; Score 7; DB 2; Length 753;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 ASAIAT 64
    |||||
Db 720 ASAIAT 726

RESULT 40
S59623
tropoelastin - sheep
C:Species: Ovis orientalis aries. Ovis ammon aries (domestic sheep)
C:Date: 23-Aug-1996 #sequence_revision 13-Mar-1997 #text_change 24-Oct-1997
C:Accession: S59623; A24758
R:Mauch, J.C.; Sandberg, L.B.; Roos, P.J.; Jimenez, F.; Christiano, A.M.; Deak, S.B.; I
  Matrix Biol. 14, 635-641, 1994
A:Title: Extensive alternate exon usage at the 5' end of the sheep tropoelastin gene.
A:Reference number: S59623
A:Accession: S59623
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-770 <MAU>
R:Yoon, K.; Davidson, J.M.; Boyd, C.; May, M.; LuValle, P.; Ornstein-Goldstein, N.; Sm:

```



Arch. Biochem. Biophys. 241, 684-691, 1985  
A>Title: Analysis of the 3' region of the sheep elastin gene.  
A;Reference number: A24758; MUID:85305763; PMID:383997  
A;Accession: A24758  
A;Molecule type: mRNA  
A;Residues: 655-669,671-716,732-770 <YOO>  
C;Superfamily: elastin  
C;Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxylysine  
F;760-765/Disulfide bonds: #status predicted

Query Match	5.0%	Score 7;	DB 2;	Length 770;
Best Local Similarity	100.0%	Pred. No. 79;		
Matches	7;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;

Qy	52	VLPGTGA 58
Db	166	VLPGTGA 172

Search completed: November 14, 2003, 10:52:29  
Job time : 38 secs

34	6	4.3	157	1	HS2_MOUSE	Q54792	mus musculus
35	6	4.3	157	1	HS2_RAT	P35429	rattus norv
36	6	4.3	160	1	YRN5_CAEEL	Q09419	caenorhabdi
37	6	4.3	162	1	GTR4_CANPA	Ox2422	canis famli
38	6	4.3	167	1	B3AR_MERUN	Ox2032	meriones un
39	6	4.3	167	1	ILVH_MYCAV	Q59499	mycobacteri
40	6	4.3	168	1	ILVH_MYCTU	O53249	mycobacteri
41	6	4.3	169	1	ILVH_MYCTU	O33113	mycobacteri
42	6	4.3	173	1	HS2_HUMAN	Q9Y543	homo sapien
43	6	4.3	176	1	PRND_HUMAN	Q9UKY0	homo sapien
44	6	4.3	176	1	RL6_HABIN	P44347	haemophilus
45	6	4.3	177	1	MOC3_MYCTU	O53376	mycobacteri
46	6	4.3	177	1	RELX_HESAU	Q84171	mesocricetu
47	6	4.3	179	1	PRND_MOUSE	Q9GUG3	mus musculus
48	6	4.3	179	1	SDPI_HUMAN	P57086	homo sapien
49	6	4.3	180	1	DADR_RABIT	O02664	oryctolagus
50	6	4.3	182	1	TRIF_CHICK	P02644	gallus galli
51	6	4.3	184	1	MCAR_MOUSE	P56450	mus musculus
52	6	4.3	184	1	TRIS_RABIT	P02645	oryctolagus
53	6	4.3	186	1	TRIS_HUMAN	P19237	homo sapien
54	6	4.3	186	1	TRIS_RAT	P19237	homo sapien
55	6	4.3	192	1	YEAB_ECOLI	P34337	escherichia
56	6	4.3	195	1	NRTN_MOUSE	P37463	mus musculus
57	6	4.3	196	1	YSP2_LEPIN	P24845	leptospira
58	6	4.3	199	1	LOLA_VIBCH	P57069	vibrio chol
59	6	4.3	199	1	YE34_ARCFU	O28838	archaeoglob
60	6	4.3	200	1	YC68_PASMU	Q8C196	pasteurella
61	6	4.3	204	1	CPCF_SYN7P	Q44116	synchococc
62	6	4.3	207	1	UCRI_CHRV1	Q31214	chromatium
63	6	4.3	208	1	UPP_CAMJE	Q9N13	campylobact
64	6	4.3	209	1	TDXH_METTH	O26262	methanobact
65	6	4.3	209	1	TDXH_METTH	Q57109	methanobact
66	6	4.3	209	1	UPP_STAAH	Q99866	staphylococ
67	6	4.3	214	1	CHI3_HUMAN	Q96K89	homo sapien
68	6	4.3	217	1	SGS3_DROSI	P13729	drosofila
69	6	4.3	220	1	TRPF_XYLFA	P76114	escherichia
70	6	4.3	221	1	YNCC_ECOLI	Q9Kqt2	vibrio chol
71	6	4.3	225	1	KCY_VIBCH	O8P8P4	xanthomonas
72	6	4.3	226	1	KCY_XANCP	P28641	pisum sativ
73	6	4.3	227	1	KCY_ECOLI	P28641	pisum sativ
74	6	4.3	227	1	KCY_SALTI	Q8Z803	salmonella
75	6	4.3	227	1	KCY_SALTY	Q8Z803	salmonella
76	6	4.3	227	1	KCY_XANAC	Q8Z803	salmonella
77	6	4.3	230	1	KCY_ERPE	Q8Z803	salmonella
78	6	4.3	231	1	BDI2_ECOLI	Q8Z803	salmonella
79	6	4.3	232	1	DHN3_PEA	Q8Z803	salmonella
80	6	4.3	233	1	YOD2_CABEL	Q8Z803	salmonella
81	6	4.3	240	1	Y116_PASMU	Q8Z803	salmonella
82	6	4.3	246	1	MT22_SHEEP	Q8Z803	salmonella
83	6	4.3	249	1	UBIG_RHILU	Q8Z803	salmonella
84	6	4.3	252	1	RL4_ARCFU	Q8Z803	salmonella
85	6	4.3	252	1	RL4_METJA	Q8Z803	salmonella
86	6	4.3	255	1	RL4_PVRAB	Q8Z803	salmonella
87	6	4.3	255	1	RL4_PVRHO	Q8Z803	salmonella
88	6	4.3	256	1	TRUA_PYRAE	Q8Z803	salmonella
89	6	4.3	257	1	PYRK_BACLL	Q8Z803	salmonella
90	6	4.3	261	1	HIS6_MYCLE	Q8Z803	salmonella
91	6	4.3	265	1	CMFP_SCHPO	Q8Z803	salmonella
92	6	4.3	266	1	SCO2_HUMAN	Q8Z803	salmonella
93	6	4.3	271	1	FPG_SHEON	Q8E8D9	shewanella
94	6	4.3	274	1	SCF_CANEA	Q06220	canis famli
95	6	4.3	275	1	FPG_NEIMA	P57014	neisseria m
96	6	4.3	275	1	FPG_NEIMA	P55044	neisseria m
97	6	4.3	275	1	LMGN_STRUN	Q54367	streptomyce
98	6	4.3	276	1	KC2D_ARATH	O81275	arabidopsis
99	6	4.3	281	1	ALF1_PYRAB	Q9V216	pyrococcus
100	6	4.3	281	1	ALF1_PYRPU	P58314	pyrococcus
101	6	4.3	281	1	ALF1_PYRHO	O57840	pyrococcus
102	6	4.3	282	1	BI0B_HELPJ	Q9ZJK8	helicobacte
103	6	4.3	282	1	BI0B_HELPJ	O25956	helicobacte
104	6	4.3	282	1	HAT1_ARATH	P46600	arabidopsis
105	6	4.3	282	1	HAT1_ARATH	Q8PX12	methanoscac
106	6	4.3	287	1	HPX1_METMA	P18623	streptomyce

107	6	4.3	292	1	EFTS_MYCPE	Q8eug8 mycoplasma	180	6	4.3	404	1	ASSY_LISMO	Q8y5h2 listeria mo
108	6	4.3	292	1	TOXR_VIBPA	Q05938 vibrio para	181	6	4.3	404	1	YBR3_YEAST	P38083 saccharomyc
109	6	4.3	293	1	VG11_BPPH2	P04333 bacterioph	182	6	4.3	406	1	RL48_ARATH	Q9sf40 arabidopsis
110	6	4.3	294	1	TSY9_ENCCU	O62584 encaphalito	183	6	4.3	407	1	RL4A_ARATH	P45691 arabidopsis
111	6	4.3	302	1	DAPA_PROMA	P49423 prochloroco	184	6	4.3	413	1	NODC_RHISN	P50357 rhizobium s
112	6	4.3	307	1	EFTS_BARBU	Q8t66 bartonella	185	6	4.3	413	1	RF1M_YEAST	P30775 saccharomyc
113	6	4.3	307	1	EFTS_BARBU	Q9xcm5 bartonella	186	6	4.3	414	1	YLPD_CAEEL	P34384 caenorhabdi
114	6	4.3	313	1	HEW3_SALTY	Q916q2 salmoneila	187	6	4.3	416	1	PXN1_XENLA	P49263 xenopus lae
115	6	4.3	313	1	A33_HUMAN	Q99795 homo sapien	188	6	4.3	417	1	PDEF_MOUSE	P97298 mus musculu
116	6	4.3	321	1	L1FA_ECOLI	P25845 escherichia	189	6	4.3	418	1	ICP1_TOBAC	P51845 nicotiana t
117	6	4.3	321	1	YC39_CVAPA	P48279 cyanophora	190	6	4.3	420	1	ISPG_BRUME	O58117 bruceella me
118	6	4.3	321	1	ASTE_ECOLI	P76215 escherichia	191	6	4.3	423	1	ARGD_KLULA	O14433 kluyveromyc
119	6	4.3	325	1	CTH1_YEAST	P47976 saccharomyc	192	6	4.3	424	1	CYAA_STIAU	P40137 stigmatella
120	6	4.3	327	1	CI30_HUMAN	Q9Y375 homo sapien	193	6	4.3	428	1	CLPX_XANAC	Q8bpi1 xanthomonas
121	6	4.3	328	1	SYFA_HELPU	Q95kf9 helicobacte	194	6	4.3	428	1	CLPX_XANCP	Q8bpy5 xanthomonas
122	6	4.3	328	1	SYFA_HELPU	P56146 helicobacte	195	6	4.3	428	1	YB01_MYCPN	P75568 mycoplasma
123	6	4.3	331	1	DMPL_PSESP	P19730 pseudomonas	196	6	4.3	432	1	TIG_THETN	Q8rc26 thermoanaer
124	6	4.3	331	1	RECH_WIGBR	O842w7 wisgleswort	197	6	4.3	433	1	HTR2_HALVA	P42258 haloarcula
125	6	4.3	331	1	YOO1_CAEEL	P34633 caenorhabdi	198	6	4.3	433	1	Y384_MYCCE	P47624 mycoplasma
126	6	4.3	332	1	MC4R_PIG	Q97504 sus scrofa	199	6	4.3	438	1	YC05_MYCPN	P75871 mycoplasma
127	6	4.3	332	1	MC4R_PIG	P70596 rattus norv	200	6	4.3	438	1	DNAA_THEMEA	P46798 thermotoga
128	6	4.3	333	1	SYFA_WIGBR	O843b6 wiggleswort	201	6	4.3	446	1	DADR_HUMAN	P42288 didelphis m
129	6	4.3	334	1	MOAA_METAC	O8tug2 methanosarc	202	6	4.3	446	1	DADR_HUMAN	P21728 homo sapien
130	6	4.3	334	1	MOAA_METMA	O8px29 methanosarc	203	6	4.3	446	1	DADR_NACMU	O77680 macaca mula
131	6	4.3	337	1	LPLA_ECOLI	P32099 escherichia	204	6	4.3	446	1	DADR_MOUSE	Q61616 mus musculu
132	6	4.3	338	1	FOSE_MOUSE	P13346 mus musculu	205	6	4.3	446	1	DADR_PIG	P50130 sus scrofa
133	6	4.3	338	1	SYFA_AQUAE	O67087 aquifex ae	206	6	4.3	446	1	DADR_RAT	P18901 rattus norv
134	6	4.3	338	1	YA39_MYCPN	P75075 mycoplasma	207	6	4.3	449	1	IFR1_MOUSE	P13182 mus musculu
135	6	4.3	339	1	SYFA_CHLPN	Q926r6 chlamydia p	208	6	4.3	449	1	IFR1_RAT	P26695 rattus norv
136	6	4.3	341	1	NA37_HABIN	P44896 haemophilus	209	6	4.3	451	1	DADR_XENLA	P42889 xenopus lae
137	6	4.3	341	1	SYFA_CHLMU	Q9pl83 chlamydia m	210	6	4.3	451	1	IFR1_HUMAN	O00458 homo sapien
138	6	4.3	341	1	SYFA_MYCCE	P47436 mycoplasma	211	6	4.3	452	1	YEFF_ECOLI	P33016 escherichia
139	6	4.3	342	1	SYFA_CHLTR	O84843 chlamydia t	212	6	4.3	454	1	MUC_MESAU	Q06337 mesocricetu
140	6	4.3	342	1	UL21_HCNWA	P16846 human cytom	213	6	4.3	454	1	TRME_ECO57	O8xb41 escherichia
141	6	4.3	343	1	GPDA_RHLO	P58142 rhizobium l	214	6	4.3	454	1	TRME_ECOLI	P25522 escherichia
142	6	4.3	343	1	PI4A_PIG	Q02745 s cmpn-n-ace	215	6	4.3	454	1	TRME_SALTI	O82n8 salmonella
143	6	4.3	344	1	PYRC_NEIMA	Q9jvd6 neisseria m	216	6	4.3	454	1	TRME_SALTY	Q8ky3 salmonella
144	6	4.3	344	1	PYRC_NEIMB	Q9kod1 neisseria m	217	6	4.3	457	1	DBDR_XENLA	P42290 xenopus lae
145	6	4.3	344	1	Y240_MYCPN	P75442 mycoplasma	218	6	4.3	458	1	YNEI_XENLA	P30640 caenorhabdi
146	6	4.3	348	1	FEN_METKA	O8txu4 methanopyru	219	6	4.3	459	1	D1DR_FUGRU	P53452 fugu rubrip
147	6	4.3	351	1	AOX_ASPNG	Q74180 aspergillus	220	6	4.3	459	1	IL7R_MOUSE	P16872 mus musculu
148	6	4.3	351	1	MRAY_METH	O26830 methanobact	221	6	4.3	460	1	MURD_ENTHR	Q07669 enterococu
149	6	4.3	353	1	CCPA_ACEXY	P37697 acetobacter	222	6	4.3	460	1	NPL2_HUMAN	Q9ulw6 homo sapien
150	6	4.3	353	1	CUPS_GALME	Q24988 galleria me	223	6	4.3	460	1	NPL2_MOUSE	P51860 mus musculu
151	6	4.3	356	1	Y4NL_RHISN	P55584 rhizobium s	224	6	4.3	463	1	DSDR_FUGRU	P53454 fugu rubrip
152	6	4.3	357	1	RUVB_STRCO	Q91291 streptomyce	225	6	4.3	463	1	DSDR_XENLA	P42291 xenopus lae
153	6	4.3	360	1	DHG_THEAC	P13203 thermoplasm	226	6	4.3	469	1	NRAM_IAPAR	P06819 influenza a
154	6	4.3	363	1	ALF_ECHMU	Q9gdp32 echinococu	227	6	4.3	471	1	VG20_BPP22	O01076 bacterioph
155	6	4.3	363	1	D1DR_CARAU	P35406 carassius a	228	6	4.3	474	1	VP61_NPVOP	O10270 orgyia pseu
156	6	4.3	368	1	RH23_SCHPO	O74803 schizosacch	229	6	4.3	480	1	SYFA_METJA	Q57911 methanococc
157	6	4.3	369	1	ID12_SULTO	Q96vw9 sulfobus	230	6	4.3	481	1	PGKH_TOBAC	Q42961 nicotiana t
158	6	4.3	371	1	QPSV_APTME	O61303 apis mellif	231	6	4.3	481	1	SYFA_THEVO	Q979u4 thermoplasm
159	6	4.3	373	1	HNZA_DROGU	Q24648 droscophila	232	6	4.3	485	1	NODC_BRAJA	P26024 bradyrhizob
160	6	4.3	373	1	YSAK_CAEEL	Q93454 caenorhabdi	233	6	4.3	486	1	QUIC_ACICA	Q13522 acinetobact
161	6	4.3	374	1	HSF5_ARATH	Q987u5 arabidopsis	234	6	4.3	486	1	YDBH_SCHPO	Q10367 schizosacch
162	6	4.3	377	1	DAPA_COILA	Q39535 coix lachry	235	6	4.3	488	1	UZIP_DROME	P10379 drosophila
163	6	4.3	379	1	NER2_MOUSE	Q9jmh3 mus musculu	236	6	4.3	490	1	C71P_ARATH	Q9stk8 arabidopsis
164	6	4.3	379	1	YO11_MOUSE	P11260 mus musculu	237	6	4.3	491	1	FIBB_HUMAN	P02675 homo sapien
165	6	4.3	380	1	DAPA_MAZE	P26259 zea mays (m	238	6	4.3	492	1	C136_MYCTU	P95099 mycobacteri
166	6	4.3	383	1	DLK_HUMAN	P80170 homo sapien	239	6	4.3	492	1	DPD2_HUMAN	Q9uhl4 homo sapien
167	6	4.3	386	1	RL4_URECA	P47800 oreochromis	240	6	4.3	495	1	GDF5_MOUSE	P43027 mus musculu
168	6	4.3	386	1	Y4NL_RHISN	P49165 urechis cau	241	6	4.3	500	1	DPD2_RAT	Q9epb1 rattus norv
169	6	4.3	387	1	GCP_RICPR	Q9zeab rickettsia	242	6	4.3	501	1	GDF5_HUMAN	P43026 mus sapien
170	6	4.3	387	1	XYLR_HABIN	P45043 haemophilus	243	6	4.3	506	1	DPD2_MOUSE	Q9et22 mus musculu
171	6	4.3	394	1	HISZ_PSEAO	Q9hum5 pseudomonas	244	6	4.3	509	1	GTR4_HUMAN	P14672 homo sapien
172	6	4.3	394	1	VATC_SCHPO	Q9hdw6 schizosacch	245	6	4.3	509	1	GTR4_MOUSE	P14142 mus musculu
173	6	4.3	395	1	CVSK_STPAU	P87131 schizosacch	246	6	4.3	509	1	GTR4_MOUSE	P19357 rattus norv
174	6	4.3	396	1	EFTU_STPAU	P42479 stigmatella	247	6	4.3	509	1	STK_HYDAT	P17713 hydra atten
175	6	4.3	396	1	WR21_ORYLA	Q9y313 oryzias lat	248	6	4.3	512	1	THRC_ASHGO	Q00663 ashbya goss
176	6	4.3	399	1	ASSY_MEITH	O27322 methanobact	249	6	4.3	514	1	PSBE_PROHO	P27200 prochloroth
177	6	4.3	400	1	B3AR_MOUSE	P25962 mus musculu	250	6	4.3	517	1	MURF_MYCLE	O69556 mycobacteri
178	6	4.3	402	1	YBJU_ECOLI	P75810 escherichia	251	6	4.3	519	1	GALI_SCHPO	Q9hud2 schizosacch
179	6	4.3	404	1	ASSY_LISIN	Q929s9 listeria in	252	6	4.3	523	1	E2BD_HUMAN	Q9hu10 homo sapien

253	6	4.3	523	1	E2BD_RABIT	P41111	oryctolagus	326	6	4.3	919	1	CAPP_COREF	Q8rdl3	corynebact
254	6	4.3	524	1	E2BD_MOUSE	Q61749	mus musculus	327	6	4.3	926	1	TAP1_SCHPO	O14261	schizosacch
255	6	4.3	524	1	E2BD_RAT	Q63186	rattus norv	328	6	4.3	942	1	K6PF_SCHPO	O42938	schizosacch
256	6	4.3	524	1	SVFA_METKA	Q8tym5	methanopyru	329	6	4.3	961	1	FGDI_HUMAN	P98174	homo sapien
257	6	4.3	525	1	RP54_RHSN	P22881	rhizobium s	330	6	4.3	980	1	POLG_LIV	P22338	loupung ill
258	6	4.3	528	1	MDLC_PSPPU	P20906	pseudomonas	331	6	4.3	988	1	M172_MOUSE	P97432	mus musculus
259	6	4.3	532	1	YABM_BACSU	P37555	bacillus su	332	6	4.3	990	1	YAA2_SCHPO	Q09796	schizosacch
260	6	4.3	539	1	HCP_CHLRE	Q8kbr4	chlorobium	333	6	4.3	998	1	VIOB_CHEVO	Q93400	chromobacte
261	6	4.3	554	1	ILVD_THEMA	Q9wz21	thermotoga	334	6	4.3	1004	1	SAL2_MOUSE	Q94x96	mus musculus
262	6	4.3	556	1	C948_SCHPO	P78750	schizosacch	335	6	4.3	1030	1	STK9_HUMAN	Q76039	homo sapien
263	6	4.3	581	1	LR15_HUMAN	Q8tf66	homo sapien	336	6	4.3	1055	1	XNKA_THENE	Q60042	thermotoga
264	6	4.3	589	1	DRYB_MOUSE	Q92188	mus musculus	337	6	4.3	1060	1	EG51_XENLA	P28025	xenopus lae
265	6	4.3	590	1	VPP_BPP2	P25479	bacterioph	338	6	4.3	1062	1	CARB_BACHD	Q9K9V9	bacillus ha
266	6	4.3	591	1	POXB_STRPN	Q54970	streptococc	339	6	4.3	1076	1	YF38_SCHPO	O13892	schizosacch
267	6	4.3	596	1	RGP1_DROME	Q9viw3	drosophila	340	6	4.3	1081	1	SPS2_CRAPL	O04933	craterostig
268	6	4.3	600	1	ABRA_PLARG	P23745	plasmodium	341	6	4.3	1091	1	NCA1_CHICK	P13590	gallus gall
269	6	4.3	604	1	ANYG_BHFOR	P07683	rhizopus or	342	6	4.3	1145	1	POL_EIAYV	P03371	equine infe
270	6	4.3	605	1	WSC4_YEAST	P38739	saccharomyc	343	6	4.3	1146	1	POL_EIAYV	P12542	equine infe
271	6	4.3	607	1	BD10_NICSY	P46942	nicotiana s	344	6	4.3	1161	1	DANA_YEAST	P47179	saccharomyc
272	6	4.3	612	1	ADFI_CANAL	P46589	candida alb	345	6	4.3	1176	1	NIR_NEUCR	P38681	neurospora
273	6	4.3	617	1	ILVD_BUCDN	Q9rq56	buchnera ap	346	6	4.3	1176	1	SLAP_BAGSH	P38537	bacillus sp
274	6	4.3	617	1	SYP_TREPA	Q83195	treponema p	347	6	4.3	1186	1	XPG_HUMAN	P28715	homo sapien
275	6	4.3	620	1	SVT_METJA	Q58597	methanococ	348	6	4.3	1186	1	XPG_XENLA	P14629	xenopus lae
276	6	4.3	624	1	PMW_PASMU	Q9clv8	pasteurella	349	6	4.3	1196	1	DPOL_HSVB	P28858	equine herp
277	6	4.3	627	1	ABP3_YEAST	Q08641	saccharomyc	350	6	4.3	1220	1	YMP3_CABEL	O10947	caenorhabdi
278	6	4.3	629	1	DRB_HUMAN	Q9y463	homo sapien	351	6	4.3	1221	1	YMP3_CABEL	O10947	caenorhabdi
279	6	4.3	630	1	INLB_LISMO	P25147	listeria mo	352	6	4.3	1222	1	YMH5_CABEL	P34472	caenorhabdi
280	6	4.3	630	1	MUC1_MOUSE	Q02496	mus musculus	353	6	4.3	1224	1	RP0D_PORPU	P51250	porphyra pu
281	6	4.3	630	1	YK01_SCHPO	O14115	schizosacch	354	6	4.3	1228	1	SLAP_BACST	P35825	bacillus st
282	6	4.3	638	1	CC45_SCHPO	O74113	schizosacch	355	6	4.3	1239	1	NNE3_MOUSE	Q01098	mus musculus
283	6	4.3	642	1	GATE_AERPE	Q9y9t6	aeropyrum p	356	6	4.3	1240	1	YQJ3_CAEEL	Q09550	caenorhabdi
284	6	4.3	645	1	SVT_STAAM	Q99th9	staphylococ	357	6	4.3	1259	1	AUT2_HUMAN	Q8wx77	homo sapien
285	6	4.3	652	1	NAK1_SCHPO	O75011	schizosacch	358	6	4.3	1271	1	TPPS_XENLA	Q90286	xenopus lae
286	6	4.3	654	1	MCPC_BACSU	P54576	bacillus su	359	6	4.3	1271	1	PER1_HUMAN	O15534	homo sapien
287	6	4.3	656	1	YK79_MYCTU	Q10687	mycobacteri	360	6	4.3	1296	1	ASAI_ENTFA	P17953	enterococcu
288	6	4.3	659	1	Y102_MYCLE	P53525	mycobacteri	361	6	4.3	1312	1	PIPI_DROME	P25455	drosophila
289	6	4.3	661	1	Y102_MYCTU	Q10897	mycobacteri	362	6	4.3	1315	1	CHAO_DROME	P12024	drosophila
290	6	4.3	680	1	OPDA_ECOLI	P27298	escherichia	363	6	4.3	1316	1	RPOC_MYCLE	P310761	mycobacteri
291	6	4.3	680	1	OPDA_SALTY	P27237	salmonella	364	6	4.3	1403	1	YGN1_YEAST	P53127	saccharomyc
292	6	4.3	683	1	APCE_SYNP6	P28035	synecococc	365	6	4.3	1500	1	GRLF_CANPA	P83509	canis fami
293	6	4.3	693	1	YD01_SCHPO	Q14286	schizosacch	366	6	4.3	1504	1	SLIT_DROME	P24014	drosophila
294	6	4.3	694	1	TKT1_RHIME	P58333	rhizobium m	367	6	4.3	1508	1	BCSC_XANAC	P58938	xanthomonas
295	6	4.3	695	1	TKT2_RHIME	P56900	rhizobium m	368	6	4.3	1513	1	GRLF_HUMAN	Q8nyr4	homo sapien
296	6	4.3	710	1	Y417_HUMAN	Q43301	homo sapien	369	6	4.3	1513	1	GRLF_RAT	P81128	rattus norv
297	6	4.3	722	1	MAZ2_ECOLI	P37330	escherichia	370	6	4.3	1526	1	MY52_SCHPO	Q9us16	schizosacch
298	6	4.3	725	1	MAZ2_PSESM	Q87272	pseudomonas	371	6	4.3	1551	1	YQ12_CAEEL	Q09449	caenorhabdi
299	6	4.3	729	1	YP65_CABEL	Q92124	caenorhabdi	372	6	4.3	1589	1	CC25_YEAST	P04821	saccharomyc
300	6	4.3	731	1	MASZ_AGR7S	Q8uj85	agrobacteri	373	6	4.3	1589	1	PHP_DROME	P39769	drosophila
301	6	4.3	740	1	NLDL_HUMAN	Q8uqcl	homo sapien	374	6	4.3	1590	1	DC13_DROME	P18171	drosophila
302	6	4.3	741	1	BSQ2_DROME	P11929	drosophila	375	6	4.3	1693	1	SAS_DROME	Q04164	drosophila
303	6	4.3	743	1	ABRA_PLAFC	P22620	plasmodium	376	6	4.3	1711	1	PTPO_RAT	Q84612	rattus norv
304	6	4.3	748	1	PIIP_SALTY	P37178	salmonella	377	6	4.3	1839	1	CYAA_SACKL	P23466	saccharomyc
305	6	4.3	755	1	Y572_CHLPN	Q927y1	chlamydia p	378	6	4.3	2204	1	RRPL_NDVB	P11205	newcastle d
306	6	4.3	760	1	SEPR_HUMAN	Q12884	homo sapien	379	6	4.3	2273	1	HPA1_YEAST	P32874	saccharomyc
307	6	4.3	761	1	SEPR_MOUSE	P97321	mus musculus	380	6	4.3	2524	1	NOTC_XENLA	P21783	xenopus lae
308	6	4.3	776	1	RTN1_HUMAN	Q16799	homo sapien	381	6	4.3	2556	1	NTC1_HUMAN	P46531	homo sapien
309	6	4.3	777	1	NTRY_AZOB	P45675	azospirillum	382	6	4.3	2670	1	IP3T_RAT	Q63269	rattus norv
310	6	4.3	779	1	PHK2_RHIME	Q922a4	rhizobium m	383	6	4.3	2671	1	IP3T_HUMAN	Q14573	homo sapien
311	6	4.3	783	1	HELS_HALNI	Q8hm66	halobacteri	384	6	4.3	2834	1	IP3R_DROME	P29993	drosophila
312	6	4.3	788	1	TRSL_HCMVA	P09695	human cytom	385	6	4.3	2911	1	FN2_HUMAN	P35536	homo sapien
313	6	4.3	789	1	K6PF_HAECO	P27665	haemochnus	386	6	4.3	3172	1	ERY3_SACER	Q33133	saccharopol
314	6	4.3	800	1	T2D4_HUMAN	Q15542	homo sapien	387	6	4.3	3178	1	Y889_CABEL	Q09624	caenorhabdi
315	6	4.3	827	1	ATCU_RHIME	Q9x5x3	rhizobium m	388	6	4.3	3301	1	CLR3_MOUSE	Q91210	mus musculus
316	6	4.3	829	1	E74A_DROME	P20105	drosophila	389	6	4.3	3312	1	CLR3_HUMAN	Q9nyq7	homo sapien
317	6	4.3	831	1	PLRL_MELEGRIS	Q10994	meleagris g	390	6	4.3	3313	1	CLR3_RAT	Q88278	rattus norv
318	6	4.3	846	1	IR51_HCMVA	P09715	human cytom	391	6	4.3	4563	1	APB_HUMAN	P04114	homo sapien
319	6	4.3	858	1	ENV_HV2RO	P04577	human immun	392	6	4.3	4910	1	MDN1_YEAST	Q12019	saccharomyc
320	6	4.3	878	1	CAPF_YERPE	Q8zae8	yersinia pe	393	6	4.3	5120	1	PCLO_CHICK	Q9pus6	gallus gall
321	6	4.3	878	1	MSH4_YEAST	P40965	saccharomyc	394	5	3.5	11	BP3_BOTIN	P30423	bothrops in	
322	6	4.3	879	1	E4L1_MOUSE	Q922h5	mus musculus	395	5	3.5	11	BP4_BOTIN	P30424	bothrops in	
323	6	4.3	881	1	E4L1_HUMAN	Q9r490	homo sapien	396	5	3.5	20	PL1_LUPLU	P83363	lupinus lut	
324	6	4.3	881	1	HELI_HSVB	P28934	equine herp	397	5	3.5	20	PL4_LUPLU	P83366	lupinus lut	
325	6	4.3	913	1	E4LB_HUMAN	Q9h329	homo sapien	398	5	3.5	20	PL6_LUPLU	P83368	lupinus lut	

399	1	DCUP_RHOSH	3.5	27	P32920 rhodobacter	472	5	3.5	89	1	VKIL_LAMB	P01758 bacterioph
400	1	MT2_ERANA	3.5	28	Q96353 brassica na	473	5	3.5	89	1	Y008_TREPA	O81053 treponema p
401	1	LPID_ECOLI	3.5	32	P03060 escherichia	474	5	3.5	89	1	Y007_SCHPO	O09678 schizosacch
402	1	PHNS_DESMU	3.5	32	P13062 desulfovibr	475	5	3.5	89	1	Y07A_ECOLI	P09339 escherichia
403	1	CHCD_ANTIPO	3.5	37	P08931 antheraea p	476	5	3.5	90	1	BAFL_DROME	Q9vlu0 drosophila
404	1	RL36_CORGL	3.5	40	Q8mm8 corynebacte	477	5	3.5	90	1	NOLS_RHIME	Q52975 rhizobium m
405	1	Y04D_BPT4	3.5	43	P07080 bacterioph	478	5	3.5	90	1	DBH_RICPR	Q9zct1 rickettsia
406	1	YC17_CVACA	3.5	43	Q9tm07 cyanidium c	479	5	3.5	91	1	DBH_LACLA	Q9c164 lactococcus
407	1	OPO2_OTICA	3.5	44	P83314 opisthoph	480	5	3.5	91	1	YFDM_ECOLI	P76509 escherichia
408	1	YC17_CVABA	3.5	49	P48367 cyanophora	481	5	3.5	92	1	C553_BACPA	P82599 bacillus pa
409	1	STPI_BOVIN	3.5	54	P17305 bos taurus	482	5	3.5	92	1	RL37_EMENI	Q960t1 emericella
410	1	STPI_SHEEP	3.5	54	P22613 ovis aries	483	5	3.5	93	1	SH3N_HUMAN	Q9n299 homo sapien
411	1	ATPB_CORCN	3.5	55	Q9td16 corythaixoi	484	5	3.5	93	1	Y1A1_RHISP	P17983 rhizobium s
412	1	ATPB_MUSVO	3.5	55	Q9tb18 musophaga v	485	5	3.5	94	1	NAG7_HUMAN	Q9y6c7 homo sapien
413	1	BLYM_HUMAN	3.5	58	P01124 homo sapien	486	5	3.5	94	1	YGBF_ECOLI	P45956 escherichia
414	1	Y088_RICPR	3.5	59	Q9ze59 rickettsia	487	5	3.5	95	1	HMVG_MOUSE	P17095 mus musculu
415	1	RZ0D_ECOLI	3.5	60	P58041 escherichia	488	5	3.5	95	1	SR19_PYRAE	Q8ztb7 pyrobaculum
416	1	UCRY_SOLTU	3.5	61	P48505 solanum tub	489	5	3.5	95	1	SR19_PYRAE	Q8ztb7 pyrobaculum
417	1	V238_FOWPV	3.5	61	Q9v216 foalpo vir	490	5	3.5	95	1	VINI_BPT4	P03718 bacterioph
418	1	LANN_STRMU	3.5	63	O68586 streptococ	491	5	3.5	95	1	YDHI_SCHPO	Q92346 schizosacch
419	1	RPCX_SCHPO	3.5	63	P48011 schizosacch	492	5	3.5	95	1	YNS9_XANAP	Q8pk18 xanthomonas
420	1	IVBI_BUNFA	3.5	65	P25660 burgarus fa	493	5	3.5	96	1	TEN3_TENMO	Q27270 tenebrio m
421	1	RL35_BUCBP	3.5	65	Q89av9 buchnera ap	494	5	3.5	96	1	Y039_CHLNP	Q929e2 chlamydia p
422	1	Y234_BACHD	3.5	65	Q9kg78 bacillus ha	495	5	3.5	96	1	YS98_MYCLE	Q33024 mycobacteri
423	1	YN43_ARCFU	3.5	65	O27941 archaeoglob	496	5	3.5	99	1	D127_HUMAN	Q9hlma4 homo sapien
424	1	MP51_MYCAV	3.5	66	O52957 mycobacteri	497	5	3.5	100	1	APA2_MACFA	P16656 macaca fasc
425	1	RS14_TOBAC	3.5	66	P93377 nicotiana t	498	5	3.5	100	1	RS14_ANASP	O8v966 anabaena sp
426	1	XEN2_XENLA	3.5	66	P38951 xenopus lae	499	5	3.5	100	1	SIS_LYNGST	P42579 lymnaea sca
427	1	XEN3_XENLA	3.5	66	P38952 xenopus lae	500	5	3.5	100	1	VP52_BPAPS	Q9t1p6 bacterioph
428	1	YSDA_ECOLI	3.5	66	P58096 escherichia	501	5	3.5	100	1	Y837_NEIMB	Q91z25 neisseria m
429	1	ATPL_BACME	3.5	70	P20603 bacillus me	502	5	3.5	101	1	IF52_CHICK	Q09121 gallus gall
430	1	MARB_ECOLI	3.5	72	P31121 escherichia	503	5	3.5	101	1	Y488_MYCPN	Q92605 mycoplasma
431	1	Y055_NPVAC	3.5	73	P41459 autographa	504	5	3.5	101	1	YQF4_CAELB	Q09269 caenorhadi
432	1	CSRA_BACSU	3.5	74	P33911 bacillus su	505	5	3.5	103	1	AMV1_HUMAN	Q99417 homo sapien
433	1	RL29_STRCO	3.5	74	Q910d2 streptomyce	506	5	3.5	103	1	AMV1_MOUSE	Q9eq93 mus musculu
434	1	Y1A1_MYCTU	3.5	74	P19771 mycobacteri	507	5	3.5	103	1	RL21_BORBU	O51719 borrelia bu
435	1	YHNV_SALTY	3.5	75	O06050 mycobacteri	508	5	3.5	103	1	Y446_NEIMA	Q9jv14 neisseria m
436	1	APAZ_MACMU	3.5	77	P40819 salmonella	509	5	3.5	104	1	H11_BOVIN	P02253 bos taurus
437	1	IF1C_CORMA	3.5	77	Q92853 coracua mula	510	5	3.5	104	1	YS76_MYCTU	Q10802 mycobacteri
438	1	MT2B_ARATH	3.5	77	Q95gm7 cornus mas	511	5	3.5	106	1	HMGI_CRIGR	Q9qxp3 cricetus
439	1	RL29_METKA	3.5	77	Q88805 arabidopsis	512	5	3.5	106	1	HMGI_HUMAN	P17096 homo sapien
440	1	RL29_MYCTU	3.5	77	Q8tx34 methanopyru	513	5	3.5	106	1	Y4VC_RHISN	Q53211 rhizobium s
441	1	RUXX_ARCFU	3.5	77	P95057 mycobacteri	514	5	3.5	107	1	I14K_TORMA	Q91499 torpedo mar
442	1	UL11_HSV6U	3.5	77	O29386 archaeoglob	515	5	3.5	107	1	YMCD_ECOLI	P75885 escherichia
443	1	ATPB_ASFAM	3.5	78	P24448 human herpe	516	5	3.5	108	1	HMGC_MOUSE	P52927 mus musculu
444	1	ATPB_STRDO	3.5	78	P00853 aspergillus	517	5	3.5	108	1	NIRD_SALTY	P40789 salmonella
445	1	NSGX_HUMAN	3.5	79	P21933 streptococ	518	5	3.5	108	1	Y1A4_MYCTU	Q50686 mycobacteri
446	1	RS18_UREPA	3.5	79	Q9uh64 homo sapien	519	5	3.5	109	1	HMGC_HUMAN	P52926 homo sapien
447	1	YH65_SYNY3	3.5	79	Q9ppt8 ureaplasma	520	5	3.5	109	1	RLA1_MAIZE	P52855 zea mays (m
448	1	Y222_METJA	3.5	79	P72913 synechocyst	521	5	3.5	109	1	YQCC_ECOLI	Q46919 escherichia
449	1	MT21_BRAJU	3.5	80	Q60259 methanococ	522	5	3.5	110	1	HDEA_ECOLI	P26604 escherichia
450	1	MT23_BRAJU	3.5	80	P56168 brassica ju	523	5	3.5	110	1	INS_PSAOB	Q62587 psammomya o
451	1	MT25_BRAJU	3.5	80	P56170 brassica ju	524	5	3.5	110	1	KV01_RABIT	P01882 cryptotlagus
452	1	RL4E_METJA	3.5	80	P54054 methanococ	525	5	3.5	110	1	VG12_BPPFI	P25132 bacterioph
453	1	RL29_MYCLE	3.5	80	Q32989 mycobacteri	526	5	3.5	110	1	YC75_RALSO	Q8xqf2 raietonia s
454	1	TRUA_SYNP6	3.5	81	O24712 synechococ	527	5	3.5	111	1	HM16_XENLA	P20912 xenopus lae
455	1	CXMB_CONMR	3.5	82	Q26443 conus maro	528	5	3.5	111	1	YE75_SHEON	Q8egw6 shewanella
456	1	Y056_NPVOP	3.5	83	O10314 orgyia pseu	529	5	3.5	112	1	PERX_PSEPU	P23103 pseudomonas
457	1	RR17_PORPU	3.5	83	P51305 porphyra pu	530	5	3.5	112	1	H2B_TRYCR	P27795 trypanosoma
458	1	VG03_BPMD2	3.5	83	Q05242 mycobacteri	531	5	3.5	112	1	THIO_MYCSM	Q30974 mycobacteri
459	1	RL1B_STRAM	3.5	84	Q95d82 mycobacteri	532	5	3.5	113	1	HYBF_KLEPN	Q9f0d4 klebsiella
460	1	XEN3_XENLA	3.5	84	Q95d82 mycobacteri	533	5	3.5	113	1	HYBF_KLEPN	Q9f0d4 klebsiella
461	1	Y576_MYCLE	3.5	84	Q95d82 mycobacteri	534	5	3.5	113	1	HYBF_KLEPN	Q9f0d4 klebsiella
462	1	Y576_MYCLE	3.5	84	Q95d82 mycobacteri	535	5	3.5	113	1	HYBF_KLEPN	Q9f0d4 klebsiella
463	1	Y576_MYCLE	3.5	84	Q95d82 mycobacteri	536	5	3.5	113	1	HYBF_KLEPN	Q9f0d4 klebsiella
464	1	Y576_MYCLE	3.5	84	Q95d82 mycobacteri	537	5	3.5	113	1	HYBF_KLEPN	Q9f0d4 klebsiella
465	1	Y576_MYCLE	3.5	84	Q95d82 mycobacteri	538	5	3.5	113	1	HYBF_KLEPN	Q9f0d4 klebsiella
466	1	Y576_MYCLE	3.5	84	Q95d82 mycobacteri	539	5	3.5	113	1	HYBF_KLEPN	Q9f0d4 klebsiella
467	1	Y576_MYCLE	3.5	84	Q95d82 mycobacteri	540	5	3.5	113	1	HYBF_KLEPN	Q9f0d4 klebsiella
468	1	Y576_MYCLE	3.5	84	Q95d82 mycobacteri	541	5	3.5	113	1	HYBF_KLEPN	Q9f0d4 klebsiella
469	1	Y576_MYCLE	3.5	84	Q95d82 mycobacteri	542	5	3.5	113	1	HYBF_KLEPN	Q9f0d4 klebsiella
470	1	Y576_MYCLE	3.5	84	Q95d82 mycobacteri	543	5	3.5	113	1	HYBF_KLEPN	Q9f0d4 klebsiella
471	1	Y576_MYCLE	3.5	84	Q95d82 mycobacteri	544	5	3.5	113	1	HYBF_KLEPN	Q9f0d4 klebsiella

545	1	ARR1	ECOLI	P15905	escherichia	618	5	3.5	134	1	RS14	TORRU	Q9xek6	tortula rur
546	1	ARR2	ECOLI	P52144	escherichia	619	5	3.5	134	1	YR1	ECOLI	P10017	escherichia
547	1	GHRL	HUMAN	Q9ubq3	homo sapien	620	5	3.5	135	1	YD39	SCHPO	Q10273	schizosacch
548	1	YGLJ	HSV8	P63344	simian herp	621	5	3.5	136	1	ACAC	PYRFU	Q51796	pyrococcus
549	1	YE26	RALS0	Q8x2g9	ralesonia s	622	5	3.5	136	1	R141	YEAST	P06367	saccharomyc
550	1	YGS8	METUA	Q59052	methanococc	623	5	3.5	137	1	NDK	STRCO	Q29277	sus scrofa
551	1	DAN3	YEAST	P38155	saccharomyc	624	5	3.5	137	1	IPPD	PIG	P50589	streptomyce
552	1	NL76	GOSHI	Q24418	gossypium h	625	5	3.5	137	1	NXT1	CAEEL	Q9u757	caenorhabdi
553	1	NLTP	WAIZE	P19656	zea mays (m	626	5	3.5	137	1	R142	YEAST	P39516	saccharomyc
554	1	PAU1	YEAST	P38924	saccharomyc	627	5	3.5	137	1	RR9	CHLVU	P56358	chlorofella v
555	1	PAU2	YEAST	P32612	saccharomyc	628	5	3.5	137	1	RS14	KLULA	P27069	kluyverichia
556	1	PAU4	YEAST	P53427	saccharomyc	629	5	3.5	137	1	Y6C6	ECOLI	Q47719	escherichia
557	1	PAU6	YEAST	P52921	saccharomyc	630	5	3.5	137	1	YUKK	BACSU	P40761	bacillus su
558	1	SPI1	HORSE	P81121	equus cabal	631	5	3.5	138	1	CXA4	PIG	Q29559	sus scrofa
559	1	YG65	YEAST	P53343	saccharomyc	632	5	3.5	138	1	RL16	MYCCE	P47404	mycoplasma
560	1	YG2F	YEAST	P53055	saccharomyc	633	5	3.5	138	1	Y29A	MYCCE	Q92b76	mycoplasma
561	1	YH56	YEAST	P38725	saccharomyc	634	5	3.5	139	1	ATPE	DICDH	P30159	dictyocota di
562	1	YQJX	BACSU	P24809	bacillus su	635	5	3.5	139	1	GOS9	ORYSA	P27349	oryza sativ
563	1	H1A	PLADU	P68894	platynereis	636	5	3.5	139	1	NUSA	HALMO	P15738	halococcus
564	1	RS11	UREPA	Q9Pqn5	ureaplasma	637	5	3.5	139	1	PLAS	ANASO	O52830	anabaena sp
565	1	SR14	ARATH	O04421	arabidopsis	638	5	3.5	139	1	PLAS	ANASO	P46444	anabaena sp
566	1	VG38	HSV11	Q00144	ictaluriid h	639	5	3.5	139	1	PLAS	ANASO	P00301	anabaena va
567	1	IN17	HUMAN	Q43005	homo sapien	640	5	3.5	139	1	RS14	SCHPO	O14129	schizosacch
568	1	RL18	HEMA	Q92ae3	thermotoga	641	5	3.5	139	1	Y201	RICPR	Q92dwa	rickettsia
569	1	YEDF	ECOLI	P39454	escherichia	642	5	3.5	139	1	YEL3	MYCPN	Q92dwa	mycoplasma
570	1	PTHA	ECOLI	P05706	escherichia	643	5	3.5	140	1	Y337	MYCPN	P75297	mycoplasma
571	1	RL14	WAIZE	P08529	zea mays (m	644	5	3.5	141	1	HBAD	COTJA	P30892	cetrutrix co
572	1	RL14	ORYSA	P12137	oryza sativ	645	5	3.5	141	1	RL16	HELPU	Q92j80	helicobacte
573	1	RL14	TOBAC	P06382	nicotiana t	646	5	3.5	141	1	RL16	HELPU	P56041	helicobacte
574	1	RS12	BORBU	O51348	borrelia bu	647	5	3.5	142	1	Y335	PYRFU	Q8u3d3	pyrococcus
575	1	MAGS	HUMAN	P43359	homo sapien	648	5	3.5	143	1	PSB2	NICSV	Q41229	nicotiana s
576	1	PA21	BOTJA	P81243	bothrops ja	649	5	3.5	143	1	PTMA	STACA	P17876	straphylococ
577	1	RS12	TREPA	O83271	treponema p	650	5	3.5	143	1	RLK3	GUITH	O46915	guillardia
578	1	YF3F	YEAST	O64212	mycobacteri	651	5	3.5	143	1	SSRP	DEIRA	Q92ru1	gainococcus
579	1	YF3F	YEAST	P31900	saccharomyc	652	5	3.5	143	1	Y763	PYRHO	O58497	pyrococcus
580	1	SY02	RABIT	P28292	oryctolagus	653	5	3.5	143	1	RS14	TRYBB	P19800	trypanosoma
581	1	GP48	RPS17	O48402	bacterioph	654	5	3.5	144	1	Y4HN	RHISH	P55441	rhizobium s
582	1	RBFA	TREPA	O83860	treponema p	655	5	3.5	145	1	CAD3	PIG	O18926	sus scrofa
583	1	YF04	MYCPN	P75282	mycoplasma	656	5	3.5	145	1	DH11	GOSHI	P09442	gossypium h
584	1	YGM1	YEAST	P53130	saccharomyc	657	5	3.5	145	1	H2A1	WHEAT	P02275	trititum ae
585	1	SECE	SCOLI	P49550	odontella s	658	5	3.5	145	1	OM20	HUMAN	Q53388	homo sapien
586	1	SECE	SCOLI	P16320	escherichia	659	5	3.5	145	1	OM20	RAT	Q62760	rattus norv
587	1	SECE	SALTY	Q919k1	salmonella	660	5	3.5	145	1	PA21	LATLA	P19000	laticauda l
588	1	YFFN	ECOLI	P76545	escherichia	661	5	3.5	145	1	PA22	LATLA	Q8uu14	laticauda l
589	1	YH91	AQUAE	O67660	aquifex aeo	662	5	3.5	145	1	PA23	LATLA	Q8uu12	laticauda l
590	1	CRB2	HALNI	O9hnm1	halobacteri	663	5	3.5	145	1	PA24	LATLA	Q8uu12	laticauda l
591	1	RS12	AQUAE	O70089	aquifex aeo	664	5	3.5	145	1	PA25	LATCO	Q8uu12	laticauda l
592	1	Y453	ARCFU	Q29796	archaeoglob	665	5	3.5	145	1	PA25	LATLA	Q8uu12	laticauda l
593	1	COAT	BPFS2	P03611	bacterioph	666	5	3.5	145	1	PA25	LATLA	Q8uu12	laticauda l
594	1	COAT	BPFS2	P03612	bacterioph	667	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
595	1	COAT	BPFS2	P03613	bacterioph	668	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
596	1	COAT	BPFS2	P03613	bacterioph	669	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
597	1	COAT	BPFS2	P03613	bacterioph	670	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
598	1	COAT	BPFS2	P03613	bacterioph	671	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
599	1	COAT	BPFS2	P03613	bacterioph	672	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
600	1	COAT	BPFS2	P03613	bacterioph	673	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
601	1	COAT	BPFS2	P03613	bacterioph	674	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
602	1	COAT	BPFS2	P03613	bacterioph	675	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
603	1	COAT	BPFS2	P03613	bacterioph	676	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
604	1	COAT	BPFS2	P03613	bacterioph	677	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
605	1	COAT	BPFS2	P03613	bacterioph	678	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
606	1	COAT	BPFS2	P03613	bacterioph	679	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
607	1	COAT	BPFS2	P03613	bacterioph	680	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
608	1	COAT	BPFS2	P03613	bacterioph	681	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
609	1	COAT	BPFS2	P03613	bacterioph	682	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
610	1	COAT	BPFS2	P03613	bacterioph	683	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
611	1	COAT	BPFS2	P03613	bacterioph	684	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
612	1	COAT	BPFS2	P03613	bacterioph	685	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
613	1	COAT	BPFS2	P03613	bacterioph	686	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
614	1	COAT	BPFS2	P03613	bacterioph	687	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
615	1	COAT	BPFS2	P03613	bacterioph	688	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
616	1	COAT	BPFS2	P03613	bacterioph	689	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l
617	1	COAT	BPFS2	P03613	bacterioph	690	5	3.5	145	1	PA25	LATSE	Q8uu12	laticauda l

691	1	RL13	CHLPN	Q928t7	chlamydia p	764	5	3.5	158	1	ATPB	SCHGA	Q07233	schizaphis
692	1	RL9	VIBCH	Q9xuy9	vibrio chol	765	5	3.5	158	1	HLB2	ARATH	O24521	arabidopsis
693	1	RL9	VIBVU	Q8dcl3	vibrio vuln	766	5	3.5	158	1	KAB3	OLDFA	P58455	oldenlandia
694	1	RL9	XANAC	Q8gm12	xanthomonas	767	5	3.5	158	1	REG3	PYRAB	Q9v2d7	pyrococcus
695	1	RL9	XANCP	Q8pac0	xanthomonas	768	5	3.5	158	1	WID	BACSU	P45910	bacillus su
696	1	RL9	XVIFA	Q8caf9	xylella fas	769	5	3.5	159	1	BV1B	BETVE	P45431	betula verr
697	1	SP17	MOUSE	Q82522	mus musculus	770	5	3.5	159	1	BV1C	BETVE	P43176	betula verr
698	1	YBAD	ECOLI	P25538	escherichia	771	5	3.5	159	1	BV1M	BETVE	P43184	betula verr
699	1	YBAD	SALTY	Q8xrd5	salmonella	772	5	3.5	159	1	BV1M	BETVE	P43186	betula verr
700	1	YBAD	SHIFL	O51824	shigella fl	773	5	3.5	159	1	HLB2	GOSHI	Q93y92	gossypium h
701	1	AROQ	BUCAI	P57479	buchnera ap	774	5	3.5	159	1	MOAC	SHEON	Q8e942	shewanella
702	1	RL14	ARATH	Q9s1h0	arabidopsis	775	5	3.5	159	1	SSRP	RHILO	Q985b9	rhizobium l
703	1	RL14	ARATH	Q9cax6	arabidopsis	776	5	3.5	159	1	SSRP	RHIME	Q92r54	rhizobium m
704	1	RL14	MAIZE	P19951	zea mays (m	777	5	3.5	159	1	TRBH	AGRTU	P54916	agrobacteri
705	1	RL14	ARATH	P42036	arabidopsis	778	5	3.5	159	1	YC36	CYAPA	P48276	cyanophora
706	1	RL19	PRHO	O59041	pyrococcus	779	5	3.5	160	1	MENG	DEIRA	Q95w10	deinococcus
707	1	RL9	VIBPA	Q8r175	vibrio para	780	5	3.5	160	1	NIA	LOTTE	P39882	lotus tetra
708	1	RS14	LUPLU	O22584	lupinus lut	781	5	3.5	160	1	RS6	UREPA	Q9ppte6	ureaplasma
709	1	YEAL	THEMA	P56944	thermotoga	782	5	3.5	160	1	SSRP	AGRT5	Q8ug12	agrobacteri
710	1	SSRP	HAEM	P44110	haemophilus	783	5	3.5	161	1	HCRC	THAAR	Q33818	thaueria aro
711	1	YP59	YEAST	Q02784	saccharomyc	784	5	3.5	161	1	OBP2	WANSE	P31419	manduca sex
712	1	DRPD	CRAPL	P22241	craterostig	785	5	3.5	161	1	PHAB	SYNEL	P50031	synecococc
713	1	NRDI	MYCGA	Q9xc21	mycoplasma	786	5	3.5	161	1	TAT8	STRCO	Q9fbk8	streptomyce
714	1	RL19	SULSO	Q9x89	sulfolobus	787	5	3.5	161	1	YA49	METJA	O59449	methanococc
715	1	RS13	AGABI	P78571	agarcicus bi	788	5	3.5	162	1	COAT	SHRV	P03581	sunh-nemp m
716	1	RS14	DROMA	P14130	drosophila	789	5	3.5	162	1	MAFG	HUMAN	O15525	homo sapien
717	1	RS14	HUMAN	P06366	homo sapien	790	5	3.5	162	1	MAFG	MOUSE	O54790	mus musculu
718	1	RS14	PODCA	Q08699	podocoryne	791	5	3.5	162	1	RRPO	LVX	P27327	lilly virus
719	1	RS14	PROCL	P48855	procambartus	792	5	3.5	162	1	SNX3	YEAST	Q08826	saccharomyc
720	1	RS14	RAT	P13471	rattus norv	793	5	3.5	162	1	YB94	SCHPO	O43003	schizosacch
721	1	SP17	HUMAN	L15506	homo sapien	794	5	3.5	163	1	RM23	YEAST	Q12487	saccharomyc
722	1	SP17	MACFA	Q19021	macaca fasc	795	5	3.5	163	1	SP17	PAPHA	Q95230	papio hamad
723	1	MGSA	SALTY	Q15007	homo sapien	796	5	3.5	163	1	YB39	MYCPN	P75259	mycoplasma
724	1	MRA2	PASMU	Q8xer5	salmonella	797	5	3.5	164	1	IMI7	SCHPO	P87130	schizosacch
725	1	MRA2	SHVU	Q9cpb5	pasteurella	798	5	3.5	165	1	COAD	CHLTE	Q8kds9	chlorobium
726	1	MRA2	SHVU	Q9f91n	shewanella	799	5	3.5	165	1	YSCH	YEREN	Q01249	yersinia en
727	1	PA2H	LATSE	Q81f92	laticauda s	800	5	3.5	165	1	YSCY	YERPE	Q00929	yersinia pe
728	1	RS14	CABEL	Q48150	caenorhabdi	801	5	3.5	166	1	IPPI	RABIT	P01099	oryctolagus
729	1	RS14	NEUCR	P19115	neurospora	802	5	3.5	166	1	PV11	BPRPD	P27382	bacterioph
730	1	SOXR	SALTY	O56144	salmonella	803	5	3.5	167	1	BQCP	CHLPN	Q92901	chlamydia p
731	1	TCPS	VIBCH	P29484	vibrio chol	804	5	3.5	167	1	FTNB	ECOLI	P53091	escherichia
732	1	Y38A	MYCPN	P75213	mycoplasma	805	5	3.5	168	1	DYR	LACIA	Q59487	lactococcus
733	1	ECPC	ELICO	Q07564	eikenella c	806	5	3.5	168	1	PSA	DICDI	P12729	dictyostell
734	1	IFSA	HUMAN	P10159	homo sapien	807	5	3.5	168	1	YF24	MYCPN	P75254	mycoplasma
735	1	IFSA	RABIT	P10160	oryctolagus	808	5	3.5	170	1	CBP4	YEAST	P37267	saccharomyc
736	1	RECX	PSEAE	P37860	pseudomonas	809	5	3.5	170	1	PLAS	LYCEE	P17340	lycopersico
737	1	RISC	ACFCU	O28856	archaeoglob	810	5	3.5	170	1	PYRR	PSEAE	Q9x6w6	pseudomonas
738	1	RS14	CHLRE	P46295	chlamydomon	811	5	3.5	171	1	IM7A	HUMAN	Q99595	homo sapien
739	1	SP17	MACCU	O62770	macropus eu	812	5	3.5	171	1	IM7A	RAT	O35092	rattus norv
740	1	Y239	LISIN	Q92567	listeria in	813	5	3.5	172	1	CBR	DUNBA	P27516	dunaliella
741	1	YUGB	HAEM	P44831	haemophilus	814	5	3.5	172	1	DEF	BORBU	O51092	borrelia bu
742	1	CYCB	CHRVU	P00154	chromatium	815	5	3.5	172	1	GSPH	PSEAE	Q00515	pseudomonas
743	1	PA26	LATCO	Q8uuu7	laticauda c	816	5	3.5	172	1	Y941	TREPA	O83911	treponema p
744	1	SOXR	ECOLI	P22538	escherichia	817	5	3.5	173	1	CF66	MOUSE	Q9d1h6	mus musculu
745	1	SSRP	TREPA	O83214	treponema p	818	5	3.5	173	1	CHB2	LYMDI	P50604	lymantria d
746	1	Y123	ARCFU	O28452	archaeoglob	819	5	3.5	173	1	NADM	METTI	Q9uxn8	methanolobu
747	1	YN85	CABEL	P34598	caenorhabdi	820	5	3.5	173	1	RL10	SVNV3	P23350	synecocyst
748	1	NUSB	RICPR	Q92e01	rickettsia	821	5	3.5	173	1	ATP7	SCHPO	O94390	schizosacch
749	1	PA23	LATCO	Q8uuu0	laticauda c	822	5	3.5	174	1	EGD2	YEAST	P38879	saccharomyc
750	1	PA23	LATCO	Q8uuu9	laticauda c	823	5	3.5	174	1	HSCB	HAELN	Q57006	haemophilus
751	1	PR13	PETCR	P19418	petroselinu	824	5	3.5	174	1	IPYR	THETH	P38576	thermus the
752	1	RL19	ARPER	Q9y1f3	aeropyrum p	825	5	3.5	174	1	RS5	SYNEL	P59126	synecococc
753	1	RL30	PYRAB	Q9v1v6	pyrococcus	826	5	3.5	174	1	Y110	MYCLE	Q9cd99	mycobacteri
754	1	RL30	PRHO	O59440	pyrococcus	827	5	3.5	174	1	Y110	MYCTU	P72046	archaeoglob
755	1	HLB2	LYCES	Q941p9	lycopersico	828	5	3.5	174	1	YB98	ARCFU	O28784	archaeoglob
756	1	NUSB	RICCN	Q92165	rickettsia	829	5	3.5	174	1	YGP3	YEAST	P53102	saccharomyc
757	1	PSPN	MOUSE	O70300	mus musculus	830	5	3.5	175	1	PP1B	STRCO	P77949	streptomyce
758	1	RS7	TREPA	O83272	treponema p	831	5	3.5	175	1	PYRE	HALNI	Q9hng2	halobacteri
759	1	YC95	AQUAE	O67324	aquifex aeo	832	5	3.5	175	1	SIVA	HUMAN	O15304	homo sapien
760	1	AROQ	BRUME	Q8y5u6	brucella me	833	5	3.5	175	1	Y240	METJA	Q57692	methanococc
761	1	IF51	SCHPO	P56289	schizosacch	834	5	3.5	176	1	MCH	NETME	Q9xpw5	methylophil
762	1	IF52	SCHPO	Q9ust4	schizosacch	835	5	3.5	176	1	VG03	BPT4	P13331	bacterioph
763	1	RBFA	BIFLO	Q8g3y4	bifidobacte	836	5	3.5	176	1	YE45	METJA	Q58840	methanococc

5	3.5	177	1	DSBB_HAEIN	P44707 haemophilus	910	5	3.5	133	1	RLPB_ECOLI	P10101 escherichia
5	3.5	177	1	LB33_ARATH	Q91h8 arabidopsis	911	5	3.5	193	1	SYFB_ERWCH	P77984 erwinia chr
5	3.5	177	1	NRFH_WOLSU	Q91e6 wolfinella s	912	5	3.5	193	1	TRPF_STRMU	Q8dvf4 streptococc
5	3.5	177	1	RL6_METTH	Q36127 methanobact	913	5	3.5	193	1	VCO7_ADEO4	Q96831 human adeno
5	3.5	178	1	KCV_METJA	Q58071 methanococc	914	5	3.5	194	1	GRPE_BACHD	Q9kd73 bacillus ha
5	3.5	179	1	ADAB_BACSU	P19220 bacillus su	915	5	3.5	194	1	H1_SALTR	Q02254 salmo trutt
5	3.5	179	1	CAS2_RAT	P02667 rattus norv	916	5	3.5	194	1	HIS7_SULTO	Q970z1 sulfolobus
5	3.5	179	1	PVPE_MYCLE	Q9cb28 mycobacteri	917	5	3.5	194	1	IPPD_MOUSE	Q60829 mus musculu
5	3.5	179	1	RUVF_DEIRA	Q9rx55 deinococcus	918	5	3.5	194	1	RL6_SCHPO	P79071 schizosacch
5	3.5	179	1	SPA7_MONDO	Q82771 monodelphis	919	5	3.5	194	1	TRPG_HELPU	Q9z1u6 helicobacte
5	3.5	180	1	SMX_CHLPN	Q9z7k6 chlamydia p	920	5	3.5	195	1	NUGM_MARPO	P14944 marchantia
5	3.5	180	1	RRF_CRICTU	P41228 cricetulus	921	5	3.5	196	1	AMIR_PSEAE	P10932 pseudomonas
5	3.5	180	1	Y52L_PRONA	Q51893 prochloroco	922	5	3.5	196	1	KTHY_ARCFU	P03175 archaeoglob
5	3.5	181	1	AG19_COCPO	Q00398 coccidioid	923	5	3.5	196	1	NODA_RHILV	P04238 rhizobium l
5	3.5	181	1	RESA_BACSU	P35160 bacillus su	924	5	3.5	196	1	NUDH_HAEIN	P57045 haemophilus
5	3.5	181	1	YG34_MYCPN	P75163 mycoplasma	925	5	3.5	196	1	R15E_HALNI	Q9hal2 haemobacteri
5	3.5	181	1	YUAB_BACSU	P71014 bacillus su	926	5	3.5	196	1	R18A_HUMAN	Q9rvs2 homo sapien
5	3.5	182	1	INB_RAT	Q97209 borellia bu	927	5	3.5	196	1	RK3_ODOSI	P49569 odontella s
5	3.5	182	1	JDP_PROMO	Q9cvp3 drosophila	928	5	3.5	196	1	WBGJ_ECOLI	P16011 bacteriopho
5	3.5	182	1	ORN_HAEIN	P45340 haemophilus	929	5	3.5	196	1	WBBJ_ECOLI	P37750 escherichia
5	3.5	183	1	AACI_DICDI	P41195 dictyosteli	930	5	3.5	196	1	YCHG_ECOLI	P30192 escherichia
5	3.5	183	1	CITX_ECOLI	Q9fk03 escherichia	931	5	3.5	196	1	YJVB_YEAST	P40892 saccharomyc
5	3.5	183	1	CITX_ECOLI	P77563 escherichia	932	5	3.5	197	1	DCTR_RHOCA	P37740 rhodobacter
5	3.5	183	1	YEG2_ARCFU	Q28810 archaeoglob	933	5	3.5	197	1	DHNI_PEA	P28639 pisum sativ
5	3.5	183	1	YGJV_ECOLI	P42603 escherichia	934	5	3.5	197	1	HIS2_THEMA	Q9x0c5 thermotoga
5	3.5	184	1	YB39_WARPO	P38477 marchantia	935	5	3.5	197	1	LEXA_THENE	Q86948 thermotoga
5	3.5	184	1	INB_RAT	P70499 rattus norv	936	5	3.5	197	1	NRTH_HUMAN	Q99748 homo sapien
5	3.5	184	1	RB55_ACEBL	P16133 acetabulari	937	5	3.5	197	1	NUOC_RICPR	Q9zdh3 rickettsia
5	3.5	184	1	KY17_YEAST	Q02205 saccharomyc	938	5	3.5	197	1	RECR_XANAC	Q8p9g0 xanthomonas
5	3.5	185	1	COMI_DICDI	Q03380 dictyosteli	939	5	3.5	197	1	RECR_XANCP	Q8pbw4 xanthomonas
5	3.5	185	1	RPEC_MYCJE	Q10679 mycobacteri	940	5	3.5	197	1	Y057_METKA	Q8t282 methanopyru
5	3.5	185	1	RRF_CAMYE	Q9pi10 campylobact	941	5	3.5	198	1	SOC2_MOUSE	Q135717 mus musculu
5	3.5	185	1	RRF_CLOPE	Q46293 clostridium	942	5	3.5	198	1	SOC2_RAT	Q88582 rattus norv
5	3.5	185	1	Y25E_AERPE	Q9yf42 aeropyrum p	943	5	3.5	198	1	F93_YERPE	P58635 yerseinia pe
5	3.5	186	1	ATPD_RHOSH	P05438 rhodospiril	944	5	3.5	198	1	YGSM_YEAST	P40325 saccharomyc
5	3.5	186	1	EPF_NEIMA	Q9jnu2 neisseria m	945	5	3.5	199	1	RECO_HUMAN	P35243 homo sapien
5	3.5	186	1	EPF_NEIMA	Q9jzq8 neisseria m	946	5	3.5	199	1	CD69_HUMAN	P07108 homo sapien
5	3.5	186	1	SODC_NEIMA	P57005 neisseria m	947	5	3.5	199	1	RT14_CAEEL	P49391 caenorhabdi
5	3.5	186	1	SODC_NEIMB	Q59623 neisseria m	948	5	3.5	199	1	Y002_RHIME	Q92tfl1 rhizobium m
5	3.5	186	1	TPBE_HALNY	Q9hn56 halobacteri	949	5	3.5	199	1	YA70_DEIRA	Q9rvf9 deinococcus
5	3.5	186	1	YCEB_SALTY	P40882 salmonella	950	5	3.5	200	1	LEUD_CAMJE	Q9plw2 campylobact
5	3.5	186	1	YJ41_PYRHO	Q59604 pyrococcus	951	5	3.5	200	1	PHNB_PSEAE	P09786 pseudomonas
5	3.5	187	1	KIP2_HUMAN	O75838 homo sapien	952	5	3.5	200	1	REP2_ZIGFE	P13742 zygosacchar
5	3.5	187	1	KIP2_MOUSE	Q9z309 mus musculu	953	5	3.5	200	1	REP2_RAT	Q9epj3 rattus norv
5	3.5	187	1	KIP2_PYRAE	Q9z309 mus musculu	954	5	3.5	200	1	YA62_NEIMB	Q9jy99 neisseria m
5	3.5	188	1	APT_NEIMA	Q9jty4 pyrobaculum	955	5	3.5	200	1	YC61_NEIMA	Q9julu4 neisseria m
5	3.5	188	1	APT_NEIMB	Q9jty4 pyrobaculum	956	5	3.5	201	1	CUTI_MONFR	Q8tgb8 monilia f
5	3.5	188	1	GCHI_BACHD	Q9jyb4 neisseria m	957	5	3.5	201	1	Y2016_CAEEL	P52014 caenorhabdi
5	3.5	188	1	ISPZ_CAUCR	Q9kcc7 bacillus ha	958	5	3.5	201	1	GCHI_STRCO	Q9x813 streptomyce
5	3.5	188	1	YAAH_ECOLI	Q9a288 caulobacter	959	5	3.5	201	1	RECO_BOVIN	P21457 bos taurus
5	3.5	189	1	GCVR_ECOLI	P28695 escherichia	960	5	3.5	201	1	RECO_MOUSE	P21457 bos taurus
5	3.5	189	1	3MGH_DEIRA	P23483 escherichia	961	5	3.5	201	1	RECR_YERPE	P34057 mus musculu
5	3.5	190	1	MAUE_WETWE	Q9rsq0 deinococcus	962	5	3.5	201	1	SDC2_HUMAN	Q8zc97 yerseinia pe
5	3.5	190	1	NUGM_ARATH	Q90231 methylophil	963	5	3.5	201	1	IPPD_BOVIN	P34741 homo sapien
5	3.5	190	1	NUGM_ORISA	Q95748 arabidopsis	964	5	3.5	202	1	NLG3_MACMU	P07516 bos taurus
5	3.5	190	1	VPSU_SPP4	Q53322 oryza sativ	965	5	3.5	202	1	PTH_TREPA	Q8wmh2 macaca muli
5	3.5	190	1	YD07_SCHPO	P05460 bacteriopho	966	5	3.5	202	1	Y677_TREPA	Q83975 treponema p
5	3.5	191	1	CHB1_LYMDI	O13727 schizosacch	967	5	3.5	203	1	HIS5_VIBCH	Q83683 treponema p
5	3.5	191	1	Y645_WETJA	P43515 lymantria d	968	5	3.5	203	1	RECR_WYTCT	Q9kx0 vibrio chol
5	3.5	191	1	EPF_BORBU	Q58061 methanococc	969	5	3.5	203	1	RL15_NEUCR	Q69682 mycobacteri
5	3.5	192	1	IL18_PIG	O51232 borellia bu	970	5	3.5	203	1	RPOC_FISMU	Q8x034 neurospora
5	3.5	192	1	NUGM_BETTR	Q19073 sus scrofa	971	5	3.5	203	1	YNFI_HAEIN	P42074 fischerella
5	3.5	192	1	NUGM_BETVU	Q33994 beta triglyc	972	5	3.5	203	1	YPT1_SCHPO	P44102 haemophilus
5	3.5	192	1	NUGM_BETWE	Q37787 beta vulgar	973	5	3.5	204	1	IPPD_HUMAN	P11620 schizosacch
5	3.5	192	1	TREG_SERVA	Q94011 beta webbii	974	5	3.5	204	1	KTHY_VACCV	Q9ud71 homo sapien
5	3.5	193	1	BFPA_ECOLI	P00900 serratia ma	975	5	3.5	204	1	RECX_DEIRA	P13410 vaccinia vi
5	3.5	193	1	BFPA_ECO27	P58997 escherichia	976	5	3.5	204	1	SLI2_RALSO	Q9rus2 deinococcus
5	3.5	193	1	COAT_ACLGF	P33553 escherichia	977	5	3.5	204	1	VATD_ASRPE	Q9yf38 aeropyrum p
5	3.5	193	1	IL18_BOVIN	P27737 apple chlor	978	5	3.5	204	1	YTI6_CAEEL	Q10919 caenorhabdi
5	3.5	193	1	IL18_CANFA	Q9ctu73 bos taurus	979	5	3.5	205	1	ALAG_RAT	P02764 rattus norv
5	3.5	193	1	IL18_HORSE	Q9xso0 canis fami	980	5	3.5	205	1	COAT_TRTVC	P05072 tabacco rat
5	3.5	193	1	PYRR_WYTCT	P71807 mycobacteri	982	5	3.5	205	1	CYSR_SYNY3	P23285 saccharomyc
5	3.5	193	1				5	3.5	205	1		Q55854 synchocyst



P04892 bacterioph  
P46436 ascaris suu  
P33803 variola vir  
P71426 klebsiella  
Q9YB45 aetopyrum p  
Q9YB48 homo sapien  
Q9H28 thermoplas  
Q63805 mus musculu  
P33970 halobacteri  
P22492 homo sapien  
P06350 oncorhynch  
Q8E68 shewanella  
Q8T290 pyrococcus  
Q8T290 mus musculu  
P21350 mus caroli  
P07361 mus musculu  
Q95833 homo sapien

# ALIGNMENTS

RESULT 1  
OM22\_NEUCR STANDARD; PRT; 154 AA.  
AC Q07335;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Mitochondrial import receptor subunit TOM22 (Mitochondrial 22 kDa  
DE outer membrane protein) (MOM22 protein) (Translocase of outer membrane  
DE 22 kDa subunit)  
GN TOM-22 OR TOM22 OR MOM22.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Peizomycotina; Sordariomycetes;  
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=74A;  
RX MEDLINE=93351229; PubMed=8348615;  
RA Kiebler M., Keil P., Schneider H., van der Klei I.J., Pfanner N.,  
RA Neupert W.;  
RT "The mitochondrial receptor complex: a central role of MOM22 in  
RT mediating preprotein transfer from receptors to the general insertion  
RT pore."  
RL Cell 74:483-492(1993).  
RN [2]  
RP FUNCTION.  
RX MEDLINE=96016192; PubMed=7556061;  
RA Mayer A., Nargans F.E., Neupert W., Lill R.;  
RT "MOM22 is a receptor for mitochondrial targeting sequences and  
RT cooperates with MOM19";  
RL EMBO J. 14:4204-4211(1995).

CC -1- FUNCTION: Central component of the receptor complex responsible  
CC for the recognition and translocation of cytosolically synthesized  
CC mitochondrial preproteins. Together with TOM20 functions as the  
CC transit peptide receptor at the surface of the mitochondrion outer  
CC membrane and facilitates the movement of preproteins into the  
CC translocation pore. Docks TOM20 and TOM70 for interaction with the  
CC general TOM40 import pore (GIP) complex. May regulate the TOM  
CC machinery organization, stability and channel gating (By  
CC similarity).  
CC -1- SUBUNIT: Forms part of the receptor complex that consists of at  
CC least 8 different proteins (TOM5, TOM6, TOM7, TOM20, TOM22, TOM37,  
CC TOM40 and TOM70). Interacts with TOM20 and TOM70 (By similarity).  
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Mitochondrial  
CC outer membrane.  
CC -1- DOMAIN: Its cytoplasmic domain associates with the cytoplasmic  
CC domains of TOM20 and TOM70. Its intermembrane space domain  
CC provides a trans binding site for presequences and the single  
CC membrane anchor is required for a stable interaction between the

CC GIP complex proteins (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE TOM22 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; X71021; CAA50339.1; -.  
CC PIR; A40669; A40669.  
CC InterPro; IPR005683; Tom22.  
CC Pfam; PF04281; Tom22; 1.  
CC TIGRFAMs; TIGR00985; 3a0801s05tom22; 1.  
CC Receptor; Translocation; Transport; Protein transport; Outer membrane;  
CC Mitochondrion; Transmembrane. CYTOPLASMIC (POTENTIAL).  
CC DOMAIN 1 84  
CC TRANSMEM 85 105  
CC FT DOMAIN 106 154  
CC FT DOMAIN 20 30  
CC FT DOMAIN 154 AA; 16816 MW; C62EA2F3A7557439 CRC64;  
CC SQ SEQUENCE 154 AA; 16816 MW; C62EA2F3A7557439 CRC64;

Query Match 5.0%; Score 7; DB 1; Length 154;  
Best local similarity 100.0%; Pred. No. 8,1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 STTNFV 11  
| | | | |  
Db 69 STTNFV 75

## RESULT 2

U240\_DROME STANDARD; PRT; 203 AA.  
ID U240\_DROME  
AC Q9VH39;  
DT 15-SEP-2003 (Rel. 42, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Protein CG11722.  
GN CG11722 OR BCDNA:AT14909.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=107311132;  
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.V., Benos P.V., Berman B.P., Bhandal D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,  
RA Butris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foaier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hootin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Sider-Kiamis I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasagman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhen M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RA "The genome sequence of *Drosophila melanogaster*."  
RL Science 287:2185-2195 (2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley; Tissue=Testis;  
RX MEDLINE=22426066; PubMed=12517569;  
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,  
RA Rubin G.M., Celniker S.E.;  
RL "A *Drosophila* full-length cDNA resource."  
RT Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8 (2002).  
CC -1- SIMILARITY: Belongs to the UPF0240 family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AE003685; AAF54481.1; -;  
CC EXBL; AY070793; AAL48415.1; -;  
CC FlyBase; FBGN0017777; CG11722.  
SQ SEQUENCE 203 AA; 23705 MW; 325F385D98DC5F99 CRC64;  
Query Match 5.0%; Score 7; DB 1; Length 203;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 29 LLRVKSS 35  
DB 192 LLRVKSS 198  
RESULT 3  
ID MOTB TREPA STANDARD; PRT; 238 AA.  
AC 007887;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Chemotaxis motB protein (Motility protein B).  
GN MOTB OR TP0724.  
OS Treponema pallidum.  
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.  
OX NCBI\_TaxID=160;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96345631; PubMed=8755894;  
RA Linberger R.J., Slivinski L.L., El-Afandi M.C.T., Dantuono L.A.;  
RT "Organization, transcription, and expression of the 5' region of the  
RT fla operon of *Treponema phagogenitalis* and *Treponema pallidum*."  
RL J. Bacteriol. 178:4628-4634 (1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nichols;  
RX MEDLINE=98332770; PubMed=9665876;  
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,

RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,  
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,  
RA McDonald L., Artach P., Bowman C., Coston M.D., Fujii C., Garland S.,  
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
RA Venter J.C.;  
RT "Complete genome sequence of *Treponema pallidum*, the syphilis  
RT spirochete."  
RL Science 281:375-388 (1998).  
CC -1- FUNCTION: REQUIRED FOR THE ROTATION OF THE FLAGELLAR MOTOR. MIGHT  
CC BE A LINKER THAT FASTENS THE TORQUE-GENERATING MACHINERY TO THE  
CC CELL WALL (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.  
CC -1- SIMILARITY: BELONGS TO THE MOTB FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U28219; AAB61254.1; -;  
CC EXBL; AE001244; AAC65689.1; -;  
CC FIRM; H71291; H71291.  
CC TIGR; TP0724; -;  
CC InterPro: IPR006665; OmpA/MotB.  
CC Pfam: PF00691; OmpA; 1.  
CC ProDom; PD000930; OmpA/MotB; 1.  
CC Chemotaxis; Flagella; Transmembrane; Inner membrane;  
CC Flagellar rotation; Complete cytoplasmic (POTENTIAL).  
FT DOMAIN 1 14  
FT TRANSMEM 15 34  
FT DOMAIN 35 238  
FT PERIPLASMIC (POTENTIAL).  
SQ SEQUENCE 238 AA; 26050 MW; OAB5FCDBE9852E1F CRC64;  
Query Match 5.0%; Score 7; DB 1; Length 238;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 26 REALLRV 32  
DB 138 REALLRV 144  
RESULT 4  
ID PFG MYCTU STANDARD; PRT; 289 AA.  
AC Q10559;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Formamidopyrimidine-DNA glycosylase (EC 3.2.2.23) (FAPY-DNA  
DE glycosylase).  
GN MUTM OR PPG OR RV2924C OR MT2994 OR MTCY338.13C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;  
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Broch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornaby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;  
RT "Deciphering the biology of *Mycobacterium tuberculosis* from the

RT complete genome sequence.";  
 RL Nature 393:537-544 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=CDC 1551 / Oshkosh;  
 RC Fleischnann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kelner J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Usterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THIS ENZYME MAY PLAY A SIGNIFICANT ROLE IN PROCESSES  
 CC LEADING TO RECOVERY FROM MUTAGENESIS AND/OR CELL DEATH BY  
 CC ALKYLATING AGENTS (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY. Hydrolysis of DNA containing ring-opened N(7)-  
 CC methylguanine residues, releasing 2,6-diamino-4-hydroxy-5-(N-  
 CC methyl)formamidopyrimidine.  
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE PFG FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Z74697; CA98987.1; -;  
 CC EMBL; AE007121; AAK47321.1; -;  
 CC PIR; D70748; D70748.  
 CC HSP; O50606; 1EE8.  
 CC TIGR; MT2994; -;  
 CC TubercuList; RV2924c; -;  
 CC HAMAP; MF\_00103; -; 1.  
 CC InterPro; IPR000191; Fapy DNA glyco.  
 CC InterPro; IPR000214; Fapy DNA glyco. zn.  
 CC Pfam; PF01149; Fapy DNA glyco. 1.  
 CC ProDom; PD003680; Fapy DNA glyco. 1.  
 CC TIGRfams; TIGR00577; fpg; 1.  
 CC PROSITE; PS01242; fpg; 1.  
 CC DNA repair; Hydrolase; Glycosidase; Zinc; Zinc-finger;  
 KW Complete proteome.  
 FT ZN FING 260 283 POTENTIAL.  
 FT ZN FING 260 283  
 SQ SEQUENCE 289 AA; 31950 MW; E00B94A70DC2904E CRC64;

Query Match 5.0%; Score 7; DB 1; Length 289;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 DVMREAL 29  
 DB 219 DVMREAL 225

RESULT 5  
 ID\_NIFH\_CVAA5 STANDARD; PRT; 327 AA.  
 AC 007641;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Nitrogenase iron protein (EC 1.18.6.1) (Nitrogenase component II)  
 DE (Nitrogenase Fe protein) (Nitrogenase reductase).  
 GN NIFH.  
 OS Cyanobacteria (strain ATCC 51142).  
 OC Bacteria; Cyanobacteria; Chroococcales; Cyanotherae.  
 OX NCBI\_TaxID=43989;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=20062360; PubMed=10594374;  
 RX

RA Colon-Lopez M.S., Tang H.-Y., Tucker D.L., Sherman L.A.;  
 RT "Analysis of the nifHDK operon and structure of the NifH protein from  
 RT the unicellular, diazotrophic cyanobacterium, Cyanothece sp.  
 RL ATCC 51142.";  
 RL Biochim. Biophys. Acta 1473:363-375 (1999).  
 CC -1- FUNCTION: THE KEY ENZYMATIC REACTIONS IN NITROGEN FIXATION ARE  
 CC CATALYZED BY THE NITROGENASE COMPLEX, WHICH HAS 2 COMPONENTS: THE  
 CC IRON PROTEIN AND THE MOLYBDENUM-IRON PROTEIN.  
 CC -1- CATALYTIC ACTIVITY: 8 reduced ferredoxin + 8 H(+) + N(2) + 16 ATP  
 CC = 8 oxidized ferredoxin + 2 NH(3) + 16 ADP + 16 phosphate.  
 CC -1- COFACTOR: BINDS ONE 4FE-4S CLUSTER PER DIMER.  
 CC -1- SUBUNIT: Homodimer (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE NIFH / BCLH / CHLL FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF003336; AAB61408.1; -;  
 CC HSP; P00459; IFF6.  
 CC HAMAP; MF\_00533; atypical; 1.  
 CC InterPro; IPR005977; NifH.  
 CC InterPro; IPR000392; NitrogenaseII.  
 CC Pfam; PF00142; fer4\_NifH; 1.  
 CC PRINTS; PRO0091; NITROGENASEII.  
 CC TIGRfams; TIGR01287; nifH; 1.  
 CC PROSITE; PS00746; NIFH\_FRXC\_1; 1.  
 CC PROSITE; PS00692; NIFH\_FRXC\_2; 1.  
 CC Oxidoreductase; Nitrogen fixation; Iron-sulfur; 4Fe-4S; ATP-binding.  
 FT NP BIND 45 52 ATP (POTENTIAL).  
 FT METAL 132 132 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 166 166 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT DOMAIN 1 37 PEST-LIKE; NOT FOUND IN OTHER NIFH.  
 SQ SEQUENCE 327 AA; D26946E425517D36 CRC64;

Query Match 5.0%; Score 7; DB 1; Length 327;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 RELSRKI 94  
 DB 275 RELSRKI 281

RESULT 6  
 ID\_MC4R\_HUMAN STANDARD; PRT; 332 AA.  
 AC P32245; Q16317;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Melanocortin-4 receptor (MC4-R).  
 GN MC4R.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93315499; PubMed=8392067;  
 RA Gantz I., Miwa H., Konda Y., Shimoto Y., Tashiro T., Waston S.J.,  
 RA Delvalle J.;  
 RT "Molecular cloning, expression, and gene localization of a fourth  
 RT melanocortin receptor.";  
 RL J. Biol. Chem. 268:15174-15179 (1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95157557; PubMed=7854347;  
 RA Mountjoy K.G., Mortrud M.T., Low M.J., Simerly R.B., Cone R.D.;

RT "Localization of the melanocortin-4 receptor (MC4-R) in  
 RT neuroendocrine and autonomic control circuits in the brain."  
 RL Mol. Endocrinol. 8:1298-1308(1994).  
 RN [3]  
 RP VARIANTS OBESITY ARG-30; VAL-37; LEU-78; MET-112; TRP-165; SER-252 AND  
 RP THR-317, AND VARIANTS VAL-103 AND LEU-251.  
 RX MEDLINE-99213392; PubMed-10199800;  
 RA Hanney A., Schmidt A., Nottelmann K., Heibult O., Becker I., Ziegler A.,  
 RA Gerber G., Sina M., Gorg T., Mayer H., Stegfried W., Fichter M.,  
 RA Renschmidt H., Hebebrand J.;  
 RT "Several mutations in the melanocortin-4 receptor gene including a  
 RT nonsense and a frameshift mutation associated with dominantly  
 RT inherited obesity in humans."  
 RL J. Clin. Endocrinol. Metab. 84:1483-1486(1999).  
 RN [4]  
 RP VARIANT OBESITY SER-274.  
 RX MEDLINE-21336749; PubMed-11443223;  
 RA Mergen M., Mergen H., Ozata M., Onur R., Onur C.;  
 RT "A novel melanocortin 4 receptor (MC4R) gene mutation associated with  
 RT morbid obesity."  
 RL J. Clin. Endocrinol. Metab. 86:3448-3448(2001).  
 CC -1- FUNCTION: RECEPTOR SPECIFIC TO THE HEPTAPEPTIDE CORE COMMON TO  
 CC ADRENOCORTICOTROPIC HORMONE AND ALPHA-, BETA-, AND GAMMA-MSH.  
 CC THIS RECEPTOR IS MEDIATED BY G PROTEINS THAT STIMULATES ADENYLATE  
 CC CYCLASE.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- TISSUE SPECIFICITY: BRAIN, PLACENTAL, AND GUT TISSUES.  
 CC -1- DISEASE: Defects in MC4R are a cause of autosomal dominant  
 CC obesity [MIM:601665].  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L08603; AAA35791.1; --  
 DR EMBL; S77415; AAB33341.1; --  
 DR PIR; A57055; A57055.  
 DR Genew; HGNC:6932; MC4R.  
 DR MIN; 155541; --  
 DR MIN; 601665; --  
 DR GO; GO:0016021; C: integral to membrane; TAS.  
 DR GO; GO:0005886; C: plasma membrane; TAS.  
 DR GO; GO:0004977; P: melanocortin receptor activity; TAS.  
 DR GO; GO:0007631; P: feeding behavior; TAS.  
 DR GO; GO:0007188; P: G-protein signaling, coupled to cAMP nucleo. . . ; TAS.  
 DR GO; GO:0007582; P: physiological processes; TAS.  
 DR InterPro; IPR000276; GPCR\_Rhodopsin.  
 DR Pfam; PF00001; 7tm.1; 1.  
 DR PRINTS; PR00237; GPCRHHODOPS.  
 DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL 1; 1.  
 DR PROSITE; PS00262; G-PROTEIN RECEPTOR FL 2; 1.  
 KW G-protein coupled receptor; Transmembrane; Glycoprotein;  
 KW phosphorylation; Lipoprotein; Palmitate; Polymorphism;  
 KW Disease mutation; Obesity.  
 FT DOMAIN 1 43 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 44 69 1 (POTENTIAL).  
 FT DOMAIN 70 81 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 82 106 2 (POTENTIAL).  
 FT DOMAIN 107 123 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 124 145 3 (POTENTIAL).  
 FT DOMAIN 146 165 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 166 186 4 (POTENTIAL).  
 FT DOMAIN 187 191 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 192 215 5 (POTENTIAL).  
 FT DOMAIN 216 248 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 249 271 6 (POTENTIAL).  
 FT DOMAIN 272 280 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 281 304 7 (POTENTIAL).

FT DOMAIN 305 332 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 3 3 N-LINKED (GLCNAC. . . ) (POTENTIAL).  
 FT CARBOHYD 17 17 N-LINKED (GLCNAC. . . ) (POTENTIAL).  
 FT CARBOHYD 26 26 N-LINKED (GLCNAC. . . ) (POTENTIAL).  
 FT LIPID 318 318 PALMITATE (POTENTIAL).  
 FT VARIANT 30 30 S -> R (IN OBESITY).  
 FT VARIANT 37 37 D -> V (IN OBESITY).  
 FT VARIANT 78 78 P -> L (IN OBESITY).  
 FT VARIANT 103 103 T -> M (IN OBESITY).  
 FT VARIANT 112 112 R -> W (IN OBESITY).  
 FT VARIANT 165 165 I -> L.  
 FT VARIANT 251 251 /FTID=VAR\_010706.  
 FT VARIANT 252 252 /FTID=VAR\_010710.  
 FT VARIANT 274 274 G -> S (IN OBESITY).  
 FT VARIANT 317 317 N -> S (IN OBESITY).  
 FT CONFLICT 169 169 I -> T (IN OBESITY).  
 FT CONFLICT 169 169 I -> S (IN REF. 2).  
 SQ SEQUENCE 332 AA; 36956 MW; 1C89752CF4DF30C CRC64;  
 Query Match 5.0%; Score 7; DB 1; Length 332;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 52 VLPQTGA 58  
 Dd 228 VLPQTGA 234  
 RESULT 7  
 SYFA\_PSEAE STANDARD; PRT; 338 AA.  
 AC Q910A3;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Phenylalanine-tRNA synthetase alpha chain (EC 6.1.1.20)  
 DE (Phenylalanine-tRNA synthetase alpha chain) (PHERS).  
 GN PHER OR PA2740.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 CX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PAO1;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Watrener P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Gabler R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.K., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Faulsen I.T.,  
 RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an  
 RT opportunistic pathogen."  
 RL Nature 406:959-964(2000).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +  
 CC diphosphate + L-phenylalanyl-tRNA(Phe).  
 CC -1- COFACTOR: Binds 2 magnesium ions per tetramer (By similarity).  
 CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (By  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.  
 CC Phe-tRNA synthetase alpha chain subfamily 1.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AE004702; AAC06128.1; -;  
 DR PIR; B83303; B83303.  
 DR HAMAP; P27001; 1PYS.  
 DR InterPro; IPR004186; Phe\_tRNA\_synth\_N.  
 DR InterPro; IPR004529; PheS.  
 DR InterPro; IPR002319; tRNA-synt 2d.  
 DR InterPro; IPR006195; tRNA\_ligase II.  
 DR Pfam; PF02912; Phe\_tRNA-synt\_N; 1.  
 DR Pfam; PF01409; tRNA-synt\_2d; 1.  
 DR TIGRFAMs; TIGR00468; PheS; 1.  
 DR PROSITE; PS00862; AA\_tRNA\_LIGASE II; 1.  
 DR AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 KW Metal-binding; Magnesium; Complete proteome.  
 FT METAL 252 252 MAGNESIUM (BY SIMILARITY).  
 SQ SEQUENCE 338 AA; 38063 MW; 6FEA3219E3322F0FC CRC64;

Query Match 5.0%; Score 7; DB 1; Length 338;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 36 ERLAMLR 42  
 Db 311 ERLAMLR 317

# RESULT 8

SYDA THETH STANDARD; PRT; 350 AA.  
 AC P27001;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Phenylalanyl-tRNA synthetase alpha chain (EC 6.1.1.20)  
 DE (Phenylalanine-tRNA ligase alpha chain) (PheRS).  
 GN PHE.  
 OS Thermus thermophilus.  
 OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;  
 OC Thermus.  
 OX NCBI\_TaxID=274;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HB8 / ATCC 27634;  
 RX MEDLINE=93083630; PubMed=1451792;  
 RA Keller B., Kast P., Hennecke H.;  
 RT "Cloning and sequence analysis of the phenylalanyl-tRNA synthetase  
 RL genes (pHeST) from Thermus thermophilus.";  
 RL FEBS Lett. 301:83-88(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-29.  
 RC STRAIN=HB8 / ATCC 27634;  
 RX MEDLINE=92375722; PubMed=1508711;  
 RA Kreutzer R., Kruft V., Bobkova E.V., Lavrik O.J., Sprinzl M.;  
 RT "Structure of the phenylalanyl-tRNA synthetase genes from Thermus  
 RL thermophilus HB8 and their expression in Escherichia coli.";  
 RL Nucleic Acids Res. 20:4173-4178(1992).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HB8 / ATCC 27634;  
 RA Lechler A., Kreutzer R.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).  
 RC STRAIN=HB8 / ATCC 27634;  
 RX MEDLINE=94257735; PubMed=8199244;

RA Mesyak L., Saftro M.;  
 RT "Phenylalanyl-tRNA synthetase from Thermus thermophilus has four  
 RL antiparallel folds of which only two are catalytically functional.";  
 RN Biochimie 75:1091-1098(1993).  
 RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).  
 RC STRAIN=HB8 / ATCC 27634;  
 RX MEDLINE=95393238; PubMed=7664121;  
 RA Mesyak L., Reshetnikova L., Goldgur Y., Delarue M., Saftro M.G.;  
 RT "Structure of phenylalanyl-tRNA synthetase from Thermus  
 RL thermophilus.";  
 RN Nat. Struct. Biol. 2:537-547(1995).  
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
 RX MEDLINE=99194858; PubMed=10092459;  
 RA Reshetnikova L., Moor N., Lavrik O., Vassilyev D.G.;  
 RT "Crystal structures of phenylalanyl-tRNA synthetase complexed with  
 RL phenylalanine and a phenylalanyl-adenylate analogue.";  
 RN J. Mol. Biol. 287:555-568(1999).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +  
 CC diphosphate + L-phenylalanyl-tRNA(Phe).  
 CC -!- COFACTOR: Binds 2 magnesium ions per tetramer.  
 CC -!- SUBUNIT: Tetramer of two alpha and two beta chains.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.  
 CC Phe-tRNA synthetase alpha chain subfamily 1.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; Z12118; CAA78104.1; -;  
 DR EMBL; X65609; CAA46559.1; -;  
 DR EMBL; Y15464; CAA75644.1; -;  
 DR PDB; 1PYS; 19-NOV-97.  
 DR PDB; 1B70; 09-FEB-00.  
 DR PDB; 1B7Y; 26-JAN-00.  
 DR PDB; 1E1Y; 14-DEC-01.  
 DR PDB; 1JJC; 02-NOV-01.  
 DR HAVAP; MF\_00281; -; 1.  
 DR InterPro; IPR004188; Phe\_tRNA\_synth\_N.  
 DR InterPro; IPR004529; PheS.  
 DR InterPro; IPR002319; tRNA-synt 2d.  
 DR InterPro; IPR006195; tRNA\_ligase II.  
 DR Pfam; PF02912; Phe\_tRNA-synt\_N; 1.  
 DR Pfam; PF01409; tRNA-synt\_2d; 1.  
 DR TIGRFAMs; TIGR00468; PheS; 1.  
 DR PROSITE; PS00862; AA\_tRNA\_LIGASE II; 1.  
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 KW Metal-binding; Magnesium; 3D-structure.  
 FT METAL 262 262 MAGNESIUM.  
 FT TURN 88 89  
 FT STRAND 99 99  
 FT HELIX 102 115  
 FT TURN 116 118  
 FT STRAND 120 121  
 FT STRAND 127 129  
 FT HELIX 130 133  
 FT TURN 134 138  
 FT HELIX 144 148  
 FT STRAND 152 155  
 FT TURN 157 158  
 FT STRAND 162 162  
 FT TURN 164 165  
 FT STRAND 168 168  
 FT STRAND 173 175  
 FT HELIX 180 189  
 FT STRAND 194 203  
 FT STRAND 209 209

RT cerevisiae.":  
J. Biol. Chem. 267:5442-5445 (1992).  
-!- FUNCTION: PARTICIPATES IN THE REGULATION OF THE ACCUMULATION  
OF ITS OWN MRNA.  
-!- MISCELLANEOUS: THERE ARE TWO GENES FOR L4 IN YEAST.  
-!- SIMILARITY: BELONGS TO THE L4E FAMILY OF RIBOSOMAL PROTEINS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL: X76078; CAA53687.1; -  
DR EMBL: J03195; AAA34974.1; -  
DR EMBL: Z35900; CAA84973.1; -  
DR EIR: S45897; S45887.  
DR FDB: 1K5Y; 22-MAY-02.  
DR SGD: S000235; REL4A.  
DR InterPro: IPR002136; Ribosomal L4/Lie.  
DR Pfam: PF00573; Ribosomal L4; 1.  
DR PROSITE: PS00939; RIBOSOMAL LIE; 1.  
KW Ribosomal protein; RNA-binding; Acetylation; Multigene family;  
KM 3D-structure.  
FT INIT\_MET 0 0  
FT MOD\_RES 1 1 ACETYLATION.  
FT CONFLICT 37 37 V -> L (IN REF. 1).  
FT CONFLICT 143 143 K -> T (IN REF. 1).  
FT CONFLICT 156 156 E -> D (IN REF. 1).  
FT CONFLICT 223 223 G -> S (IN REF. 1).  
FT CONFLICT 240 240 G -> S (IN REF. 1).  
SQ SEQUENCE 361 AA; 38961 MW; 38272ACDADC8AA5F CRC64;  
  
Query Match 5.0%; Score 7; DB 1; Length 361;  
Best Local Similarity 100.0%; Pred.No. 18;  
Matches 7; Conservative 0; Mismatches 0; Gaps 0;  
  
QY 58 ASAAIAAT 64  
DB 122 ASAAIAAT 128  
  
RESULT 10  
RL4B YEAST  
ID RL4B YEAST STANDARD; PRT; 361 AA.  
AC P49626;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE 60S ribosomal protein L4-B (L2B) (RP2).  
GN RPL4B OR RPL2B OR YDR012W OR YDR119.17.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c / AB972;  
RA Murphy L., Richards C., Gentles S., Harris D., Barrell B.G.,  
RA Rajandream M.A.;  
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97051598; PubMed=8896275;  
RA Eide L.G., Sander C., Prydz H.;  
RT "Sequencing and analysis of a 35.4 kb region on the right arm of  
RT chromosome IV from Saccharomyces cerevisiae reveal 23 open reading  
RT frames."  
RL Yeast 12:1085-1090 (1996).  
RN [3]  
RP SEQUENCE OF 1-199 FROM N.A.

FT TURN 210 211  
FT STRAND 212 212  
FT TURN 215 225  
FT TURN 226 227  
FT TURN 230 245  
FT TURN 247 248  
FT STRAND 251 255  
FT TURN 259 260  
FT STRAND 261 271  
FT TURN 272 275  
FT STRAND 276 286  
FT TURN 288 300  
FT TURN 301 302  
FT STRAND 311 318  
FT TURN 319 327  
FT TURN 328 328  
FT TURN 332 337  
FT TURN 340 343  
FT TURN 344 346  
FT STRAND 350 AA; 39259 MW; DEB02377CD9A461 CRC64;  
SQ SEQUENCE 350 AA; 39259 MW; DEB02377CD9A461 CRC64;  
  
Query Match 5.0%; Score 7; DB 1; Length 350;  
Best Local Similarity 100.0%; Pred.No. 17;  
Matches 7; Conservative 0; Mismatches 0; Gaps 0;  
  
QY 36 ERLAMLR 42  
DB 320 ERLAMLR 326  
  
RESULT 9  
RL4A YEAST  
ID RL4A YEAST STANDARD; PRT; 361 AA.  
AC P10664;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE 60S ribosomal protein L4-A (L2A) (RP2).  
GN RPL4A OR RPL2A OR RPL2 OR YBR031W OR YBR0315.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88198160; PubMed=2834365;  
RA Presutti C., Lucicli A., Bozzoni I.;  
RT "Ribosomal protein L2 in Saccharomyces cerevisiae is homologous to  
RT ribosomal protein L1 in Xenopus laevis. Isolation and  
RT characterization of the genes."  
RL J. Biol. Chem. 263:6188-6192 (1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c;  
RX MEDLINE=94378725; PubMed=8091864;  
RA Smits P.H.M., de Haan M., Maat C., Grivell L.A.;  
RT "The complete sequence of a 33 kb fragment on the right arm of  
RT chromosome II from Saccharomyces cerevisiae reveals 16 open reading  
RT frames, including ten new open reading frames, five previously  
RT identified genes and a homologue of the SCOL gene."  
RL Yeast 10:S75-380 (1994).  
RN [3]  
RP L2 MRNA REGULATION.  
RX MEDLINE=91293097; PubMed=2065661;  
RA Presutti C., Ciafre S.-A., Bozzoni I.;  
RT "The ribosomal protein L2 in S. cerevisiae controls the level of  
RT accumulation of its own mRNA."  
RL EMBO J. 10:2215-2221 (1991).  
RN [4]  
RP SEQUENCE OF 1-20.  
RX MEDLINE=92184799; PubMed=1544921;  
RA Takakura H., Tsunasawa S., Miyagi M., Warner J.R.;  
RT "NH2-terminal acetylation of ribosomal proteins of Saccharomycetes

```
RX MEDLINE=89096952; PubMed=3062369;
RA Lucifoli A., Presutti C., Ciafre S., Caffarelli E., Fragapane P.,
RT Bozzoni I.;
RA "Gene dosage alteration of L2 ribosomal protein genes in
RT Saccharomyces cerevisiae: effects on ribosome synthesis.";
RL Mol. Cell. Biol. 8:4792-4798(1988).
RN [4].
RP SEQUENCE OF 1-20.
RX MEDLINE=92184799; PubMed=1544921;
RA Takakura H., Tsunawasa S., Miyagi M., Warner J.R.;
RA "NH2-terminal acetylation of ribosomal proteins of Saccharomyces
RT cerevisiae.";
RL J. Biol. Chem. 267:5442-5445(1992).
CC -!- FUNCTION: PARTICIPATES IN THE REGULATION OF THE ACCUMULATION
CC OF ITS OWN MRNA.
CC -!- MISCELLANEOUS: THERE ARE TWO GENES FOR L4 IN YEAST.
CC -!- SIMILARITY: BELONGS TO THE L4E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z48008; CAA88072.1; -
DR EMBL; X95966; CAA65204.1; -
DR EMBL; Z74308; CAA88832.1; -
DR EMBL; M22583; AAA34975.1; -
DR PIR; S50993; S50993.
DR SGD; S0002419; RPL4B.
DR InterPro; IPR002136; Ribosomal L4/L1e.
DR Pfam; PF00573; Ribosomal L4; 1.
DR PROSITE; PS00939; RIBOSOMAL L1E; 1.
KW Ribosomal protein; RNA-binding; Acetylation; Multigene family.
FT INIT MET 0
FT MOD_RES 1
FT CONFLICT 87
FT CONFLICT 87
SQ SEQUENCE 361 AA; 38931 MW; 38272ACD4DC8B62F CRC64;

Query Match 5.0%; Score 7; DB 1; Length 361;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 ASATAAT 64
DB 122 ASATAAT 128

RESULT 11
MK32 YEAST
AC P23060;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE MAK32 protein.
GN MAK32 OR YCR019W OR YCR19W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1].
RP SEQUENCE FROM N.A.
RX MEDLINE=89131254; PubMed=3916862;
RA Toh-E A., Sahashi Y.;
RA "The PET18 locus of Saccharomyces cerevisiae: a complex locus
RT containing multiple genes.";
RL Yeast 1:159-171(1985).
RN [2].
RP SEQUENCE FROM N.A.
RA Feldmann H., Mannhaupt G., Vetter I.;

Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: NECESSARY FOR THE STRUCTURAL STABILITY OF L-A DOUBLE-
STRANDED RNA-CONTAINING PARTICLES. NECESSARY FOR GROWTH AT 37-
DEGREES CELSIUS AS WELL AS FOR MAINTENANCE OF THE KILLER PLASMID.
-!- SIMILARITY: TO S-POMBE SPAC4G8.14C.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; X59720; CAA42310.1; -
DR PIR; S19429; BVEYK2.
DR SGD; S0000612; MAK32.
DR GO; GO:0019048; P.virus-host interaction; IMP.
DR InterPro; IPR002173; PfKB.
DR Pfam; PF00294; pfKB; 1.
FT CONFLICT 15
FT CONFLICT 82
FT CONFLICT 82
SQ SEQUENCE 363 AA; 40783 MW; CFB358F8ACF6EA4C CRC64;

Query Match 5.0%; Score 7; DB 1; Length 363;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 LLRVKSS 35
DB 158 LLRVKSS 164

RESULT 12
NODC_RHILV
ID NODC_RHILV STANDARD; PRT; 424 AA.
AC P04340;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE N-acetylglucosaminyltransferase (EC 2.4.1.-) (Modulation protein C).
GN NODC.
OS Rhizobium leguminosarum (biovar viciae).
OG Plasmid sym pRLUJI.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=387;
RN [1].
RP SEQUENCE FROM N.A.
RC STRAIN=248;
RX MEDLINE=85087952; PubMed=6514582;
RA Rossen L., Johnston A.W.B., Downie J.A.;
RT "DNA sequence of the Rhizobium leguminosarum nodulation genes nodAB
RT and C required for root hair curling.";
RL Nucleic Acids Res. 12:9497-9508(1984).
RN [2].
RP SEQUENCE OF 91-180 FROM N.A.
RC STRAIN=USDA 2478;
RX MEDLINE=95113787; PubMed=7814339;
RA Ueda T., Suga Y., Yahiro N., Matsuguchi T.;
RT "Phylogeny of Sym plasmids of rhizobia by PCR-based sequencing of a
RT nodC segment.";
RL J. Bacteriol. 177:468-472(1995).
CC -!- FUNCTION: INVOLVED IN THE SYNTHESIS OF NOD FACTOR, A SULFATED
CC N-ACYL-BETA-1,4-TETRASACCHARIDE OF N-ACETYLGLUCOSAMINE WHICH
CC INITIATES A SERIES OF EVENTS IN THE HOST PLANT SPECIES LEADING
CC EVENTUALLY TO NODULATION.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).
CC -!- SIMILARITY: BELONGS TO THE NODC/HAS FAMILY.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; Y00548; CA868619.1; -  
CC EMBL; X01650; CA825814.1; -  
CC EMBL; D28960; BAA06086.1; -  
DR PIR; A03486; ZZZRCL.  
DR InterPro; IPR004835; Fungi\_chitin\_syn.  
DR InterPro; IPR001173; Glyco\_trans\_2.  
DR Pfam; PF03142; Chitin synth 2; 1.  
DR Pfam; PF00535; Glycosyltransf 2; 1.  
KW Transferase; Glycosyltransferase; Nodulation; Membrane; Plasmid.  
FT VARIANT 97 97 V -> E (IN STRAIN USDA 2478).  
FT VARIANT 110 110 N -> K (IN STRAIN USDA 2478).  
FT VARIANT 123 123 A -> AQ (IN STRAIN USDA 2478).  
SQ SEQUENCE 424 AA; 46255 MW; 1B1D2722A51A0FCF CRC64;

Query Match 5.0%; Score 7; DB 1; Length 424;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 SAIAATV 65  
DB 259 SAIAATV 265

## RESULT 13

AM13 MYCLE  
ID AM13 MYCLE STANDARD; PRT; 468 AA.  
AC 033040;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Putative amidase amC (EC 3.5.1.4).  
GN AMIC OR MLI596 OR MLCB250.65.  
OS Mycobacterium leprae.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1769;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TN;  
RX MEDLINE=21128732; PubMed=11234002;  
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,  
RA Wheeler P.R., Honore N., Garnier T., Churchoer C., Harris D.,  
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R.M., Devlin K., Dutroy S., Feltwell T., Fraser A., Hamlin N.,  
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
RA Barrell B.G.;  
RT "Massive gene decay in the leprosy bacillus."  
RL Nature 409:1007-1011(2001).  
CC -1- CATALYTIC ACTIVITY: A monocarboxylic acid amide + H(2)O = a  
CC monocarboxylate + NH(3).  
CC -1- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.

CC -----  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; Z97369; CAB10659.1; -  
DR EMBL; AL583922; CAC30547.1; -  
DR PIR; F87108; F87108.  
DR Leproma; MLI596; -  
DR InterPro; IPR000120; Amidase.

DR Pfam; PF01425; Amidase; 1.  
DR PROSITE; PS00571; AMIDASES; 1.  
KW Hypothetical protein; Hydrolase; Complete proteome.  
SQ SEQUENCE 468 AA; 51164 MW; A8E048D702F6A234 CRC64;

Query Match 5.0%; Score 7; DB 1; Length 468;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 SPELREL 90  
DB 269 SPELREL 275

## RESULT 14

PUR9 AQUAE  
ID PUR9 AQUAE STANDARD; PRT; 506 AA.  
AC 067775;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Bifunctional purine biosynthesis protein purH [Includes:  
DE Phosphoribosylamidoimidazolecarboxamide formyltransferase (EC 2.1.2.3)  
DE (AICAR transferase); IMP cyclohydrolase (EC 3.5.4.10) (Inosinase)  
DE (IMP synthetase) (ATIC)].  
DE PURH OR AQ\_1983.  
GN Aquifex aeolicus.  
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
OX NCBI\_TaxID=63363;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VF5;  
RX MEDLINE=98196666; PubMed=9537320;  
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aufay M., Huber R.,  
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
RT "The complete genome of the hyperthermophilic bacterium Aquifex  
RT aeolicus."  
RL Nature 392:353-358(1998).  
CC -1- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + 5-amino-1-(5-  
CC phospho-D-riboyl)imidazole-4-carboxamide = tetrahydrofolate + 5-  
CC formamido-1-(5-phospho-D-riboyl)imidazole-4-carboxamide.  
CC -1- CATALYTIC ACTIVITY: IMP + H(2)O = 5-formamido-1-(5-phospho-D-  
CC ribosyl)imidazole-4-carboxamide.  
CC -1- PATHWAY: De novo purine biosynthesis; ninth step.  
CC -1- PATHWAY: De novo purine biosynthesis; tenth step.  
CC -1- DOMAIN: THE IMP CYCLOHYDROLASE ACTIVITY RESIDUES IN THE N-TERMINAL  
CC REGION (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE PURH FAMILY.

CC -----  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AB000765; AAC07734.1; -  
DR PIR; C70468; C70468.

DR HAMAP; MF\_00139; -; 1.  
DR InterPro; IPR002695; AICARFT IMPChas.  
DR InterPro; IPR004362; MGS like.  
DR Pfam; PF01808; AICARFT IMPChas; 1.  
DR Pfam; PF02142; MGS; 1.  
DR ProDom; PD004666; AICARFT IMPChas; 1.  
DR TIGRFAMs; TIGR00355; purH; 1.  
KW Purine biosynthesis; Transferase; Hydrolase; Multifunctional enzyme;  
KW Complete proteome.

Query Match 5.0%; Score 7; DB 1; Length 506;  
Best Local Similarity 100.0%; Pred. No. 24;  
SQ SEQUENCE 506 AA; 56677 MW; 46998C8B1B1F5694 CRC64;



Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 LRELSRK 93  
 Db 335 LRELSRK 341

RESULT 15  
 ID TBX2\_MOUSE STANDARD; PRT; 701 AA.  
 AC Q60707;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE T-box transcription factor TBX2 (T-box protein 2).  
 GN TBX2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND DEVELOPMENTAL EXPRESSION.  
 RC TISSUE=Embryo;  
 RX MEDLINE=95004605; PubMed=7920656;  
 RA Bollag R.J., Siegfried Z., Cebra-Thomas J.A., Garvey N., Davison E.M.,  
 RA Silver L.M.;  
 RT "An ancient family of embryonically expressed mouse genes sharing a  
 RT conserved protein motif with the T locus.";  
 RT Nat. Genet. 7:383-389(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20357305; PubMed=10770922;  
 RA Carreira S., Liu B., Goding C.R.;  
 RA Chapman D.L., Garvey N., Hancock S., Alexiou M., Agulnik S.I.,  
 RA Gibson-Brown J.J., Cebra-Thomas J., Bollag R.J., Silver L.M.,  
 RA Papaioannou V.E.;  
 RT "Expression of the T-box family genes, Tbx1-Tbx5, during early mouse  
 RT development.";  
 RL Dev. Dyn. 206:379-390(1996).  
 CC -!- FUNCTION: INVOLVED IN THE TRANSCRIPTIONAL REGULATION OF GENES  
 CC REQUIRED FOR MESODERM DIFFERENTIATION. PROBABLY PLAYS A ROLE IN  
 CC LIMB PATTERN FORMATION.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC IN HEART, KIDNEY, AND OVARY.  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSION FIRST OBSERVED AT DAY 9.5 IN  
 CC THE OTIC AND OPTIC VESICLES AND IN THE FACIAL REGION AT DAY  
 CC 12.5, EXPRESSED IN THE TRIGEMINAL GANGLIA, FACIAL REGIONS, RETINA  
 CC AND LIMB BUD MESENCHYME. IN LATER STAGES, FOUND IN EAR PINNAE,  
 CC THE MILK LINE, LUNG MESENCHYME, BODY WALL, GENITAL RIDGE AND  
 CC DEVELOPING NERVOUS SYSTEM.  
 CC -!- SIMILARITY: Contains 1 T-box domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U15566; AAC52697.1; -;  
 CC EMBL; AF244917; AAF90050.1; -;  
 CC PIR; S46458; S46458.  
 CC HSSP; P24781; 1XBR.  
 CC TRANSFAC; T04358; -;  
 CC MGD; MGI:98494; Tbx2.

DR GO: 0005667; C:transcription factor complex; IDA.  
 DR GO: 0005515; F:protein binding activity; IPI.  
 DR GO: 0016564; P:transcriptional repressor activity; IDA.  
 DR GO: 0000122; P:negative regulation of transcription from P...; IDA.  
 DR InterPro; IPR001699; TF\_T-box.  
 DR Pfam; PF00907; T-box; 1.  
 DR PRINTS; PR00937; TBOX.  
 DR SMART; SMO0425; TBOX; 1.  
 DR PROSITE; PS01283; TBOX\_1; 1.  
 DR PROSITE; PS01264; TBOX\_2; 1.  
 DR PROSITE; PS0252; TBOX\_3; 1.  
 KW Transcription regulation; DNA-binding; Nuclear protein;  
 KW Developmental protein.  
 FT DOMAIN 48 63 POLY-ALA.  
 FT DNA BIND 104 277 T-BOX.  
 FT DOMAIN 572 580 POLY-ALA.  
 FT DOMAIN 586 594 POLY-ALA.  
 SQ SEQUENCE 701 AA; 74244 MW; 8D90ED6DA32B3859 CRC64;

Query Match 5.0%; Score 7; DB 1; Length 701;  
 Best Local Similarity 100.0%; Pred No. 33;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 128 HLRSLKS 134  
 Db 78 HLRSLKS 84  
 |||||  
 |||||

RESULT 16  
 ID TBX2\_HUMAN STANDARD; PRT; 702 AA.  
 AC Q13207; Q16424;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE T-box transcription factor TBX2 (T-box protein 2).  
 GN TBX2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Fetal kidney;  
 RX MEDLINE=96015055; PubMed=8530034;  
 RA Campbell C., Goodrich K., Casey G., Beatty B.;  
 RT "Cloning and mapping of a human gene (TBX2) sharing a highly conserved  
 RT protein motif with the Drosophila omb gene.";  
 RL Genomics 28:255-260(1995).  
 RN [2]  
 RP SEQUENCE OF 152-245 FROM N.A.  
 RC TISSUE=Fetal kidney;  
 RX MEDLINE=96169568; PubMed=8597636;  
 RA Law D.J., Gebuhr T., Garvey N., Agulnik S.I., Silver L.M.;  
 RT "Identification, characterization, and localization to chromosome  
 RT 17q21-22 of the human TBX2 homolog, member of a conserved  
 RT developmental gene family.";  
 RL Mamm Genome 6:793-797(1995).  
 CC -!- FUNCTION: INVOLVED IN THE TRANSCRIPTIONAL REGULATION OF GENES  
 CC REQUIRED FOR MESODERM DIFFERENTIATION. PROBABLY PLAYS A ROLE IN  
 CC LIMB PATTERN FORMATION.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (potential).  
 CC -!- TISSUE SPECIFICITY: EXPRESSED PRIMARILY IN ADULT IN KIDNEY, LUNG,  
 CC AND PLACENTA. WEAK EXPRESSION IN HEART AND OVARY.  
 CC -!- SIMILARITY: Contains 1 T-box domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U15566; AAC52697.1; -;  
 CC EMBL; AF244917; AAF90050.1; -;  
 CC PIR; S46458; S46458.  
 CC HSSP; P24781; 1XBR.  
 CC TRANSFAC; T04358; -;  
 CC MGD; MGI:98494; Tbx2.

CC -----  
DR EMBL; U28049; AAB73861.1; -;  
DR EMBL; S81264; AAB36216.1; -;  
DR PIR; G01840; G01840.  
DR HSSP; P24781; IYBR.  
DR TRANSFAC; T04351; -;  
DR Genew; HGNC:11597; TBX2.  
DR MIM; 600747; -;  
DR InterPro; IPR001699; TF\_I-box.  
DR Pfam; PF00907; T-box\_1.  
DR PRINTS; PR00937; TBOX.  
DR SMART; SMO0425; TBOX; 1.  
DR PROSITE; PS01283; TBOX\_1; 1.  
DR PROSITE; PS01264; TBOX\_2; 1.  
DR PROSITE; PS0252; TBOX\_3; 1.  
KW Transcription regulation; DNA-binding; Nuclear protein;  
KW Developmental protein.  
FT DOMAIN 48 63 POLY-ALA.  
FT DOMAIN 104 277 T-BOX.  
FT DOMAIN 507 517 POLY-GLY.  
FT DOMAIN 571 579 POLY-ALA.  
FT DOMAIN 585 593 POLY-ALA.  
FT CONFLICT 155 155 Y -> D (IN REF. 2).  
FT CONFLICT 165 168 AGKA -> TDKT (IN REF. 2).  
SQ SEQUENCE 702 AA; 74194 MW; C6477134C69D7C2C CRC64;  
  
Query Match 5.0%; Score 7; DB 1; Length 702;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 128 HLRLSKS 134  
Db 78 HLRLSKS 84  
|||||  
-----  
RESULT 17  
PTIP ECOLI STANDARD; PRT; 748 AA.  
ID PTIP ECOLI STANDARD; PRT; 748 AA.  
AC P37177;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Phosphoenolpyruvate-protein phosphotransferase ptsp (EC 2.7.3.9)  
DE (Phosphotransferase system, enzyme I) (Enzyme I-Ntr).  
GN PTSP OR B2829.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RA MEDLINE=97426617; PubMed=9278503;  
RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE OF 656-748 FROM N.A.  
RX MEDLINE=95204358; PubMed=7896715;  
RA Gan K., Sankaran K., Williams M.G., Aldea M., Rudd K.E., Kushner S.R.,  
Wu H.C.;  
RT "The umpa gene of Escherichia coli encodes  
RT phosphatidylglycerol:poliprotein diacylglyceryl transferase (lgt)  
RT and regulates thymidylate synthase levels through translational  
RT coupling";  
RL J. Bacteriol. 177:1879-1882(1995).  
RN [3]  
RP DISCUSSION OF SEQUENCE.  
RX MEDLINE=97128775; PubMed=8973315;  
RA Reizer J., Reizer A., Merrick M.J., Plunkett G. III, Rose D.J.,

RA Saier M.H. Jr.;  
RT "Novel phosphotransferase-encoding genes revealed by analysis of the  
RT Escherichia coli genome: a chimeric gene encoding an Enzyme I  
RT homologue that possesses a putative sensory transduction domain.";  
RL Gene 181:103-108(1996).  
CC -I- FUNCTION: COULD FUNCTION IN TRANSCRIPTIONAL REGULATION OF SIGMA-54  
CC DEPENDENT OPERONS IN CONJUNCTION WITH THE NTR (PTSO) AND IIA-NTR  
CC (PTSN) PROTEINS THEREBY PROVIDING A LINK BETWEEN CARBON AND  
CC NITROGEN ASSIMILATORY PATHWAYS.  
CC -I- CATALYTIC ACTIVITY: Phosphoenolpyruvate + protein L-histidine =  
CC pyruvate + protein N(pi)-phospho-L-histidine.  
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -I- SIMILARITY: BELONGS TO THE PEP-UTILIZING ENZYMES FAMILY.  
CC -I- SIMILARITY: IN THE N-TERMINAL, TO NIFA PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U29581; AAB40476.1; -;  
DR EMBL; AE000366; AAC75868.1; -;  
DR PIR; U12289; AAA69023.1; -;  
DR EcoGene; EG12188; dtsp.  
DR InterPro; IPR003018; GAP.  
DR InterPro; IPR006318; PEP\_P trans.  
DR InterPro; IPR000121; PEP\_utilizers.  
DR Pfam; PF01590; GAP; 1.  
DR Pfam; PF00391; PEP-utilizers; 1.  
DR Pfam; PF02896; PEP-utilizers; 1.  
DR ProDom; PD000940; PEP\_utilizers; 1.  
DR SMART; SMO0065; GAP; 1.  
DR TIGRfams; TIGR01417; PTS\_I\_fam; 1.  
DR PROSITE; PS00370; PEP\_ENZYMES\_PHOS\_SITE; 1.  
DR PROSITE; PS00742; PEP\_ENZYMES\_2; 1.  
KW Phosphotransferase system; Transferase; Kinase; Sugar transport;  
KW Phosphorylation; Complete proteome.  
FT DOMAIN 1 127 NIFA-LIKE.  
FT DOMAIN 128 170 LINKER.  
FT DOMAIN 171 748 ENZYME 1 DOMAIN.  
FT MOD\_RES 356 356 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 748 AA; 83715 MW; AC7137BD0AEBBF01 CRC64;  
  
Query Match 5.0%; Score 7; DB 1; Length 748;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 39 AMLRALA 45  
Db 646 AMLRALA 652  
|||||  
-----  
RESULT 18  
ELS MOUSE STANDARD; PRT; 860 AA.  
ID ELS MOUSE STANDARD; PRT; 860 AA.  
AC P54320;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Elastin precursor (Tropoelastin).  
GN ELN  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BA1B/C; TISSUE=Lung;  
RX MEDLINE=93130069; PubMed=7829060;

RA Wydner K.S., Sechler J.L., Boyd C.D., Passmore H.C.;  
RT "Use of an intron polymorphism to localize the tropoelastin gene to  
RT mouse chromosome 5 in a region of linkage conservation with human  
RT chromosome 7";  
RL Genomics 23:125-131(1994).  
CC -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND  
CC NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.  
CC -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER  
CC INTO AN EXTENSIBLE 3D NETWORK.  
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.  
CC -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U08210; AAA80155.1; -;  
CC PIR; A55721; EAMS.  
CC GMD; MGI:95317; Eln.  
CC InterPro; IPR003979; tropoelastin.  
CC PRINTS; PR01500; TROPOELASTIN.  
CC Structural protein; Repeat; Signal; Connective tissue.  
FT SIGNAL 1 27 POTENTIAL.  
FT CHAIN 28 860 ELASTIN.  
FT DISULFID 850 855 BY SIMILARITY.  
SQ SEQUENCE 860 AA; 71955 MW; 0C0BESA8A1EDD7F1 CRC64;  
  
Query Match 5.0%; Score 7; DB 1; Length 860;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 52 VLPGTGA 58  
DB 215 VLPGTGA 221

CC  
CC NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.  
CC -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER  
CC INTO AN EXTENSIBLE 3D NETWORK.  
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=8;  
CC Comment=Experimental confirmation may be lacking for some  
CC isoforms;  
CC Name=1;  
CC IsoId=Q99372-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q99372-2; Sequence=VSP\_004244;  
CC Name=3;  
CC IsoId=Q99372-3; Sequence=VSP\_004245;  
CC Name=4;  
CC IsoId=Q99372-4; Sequence=VSP\_004246;  
CC Name=5;  
CC IsoId=Q99372-5; Sequence=VSP\_004244, VSP\_004245;  
CC Name=6;  
CC IsoId=Q99372-6; Sequence=VSP\_004245, VSP\_004246;  
CC Name=7;  
CC IsoId=Q99372-7; Sequence=VSP\_004244, VSP\_004246;  
CC Name=8;  
CC IsoId=Q99372-8; Sequence=VSP\_004244, VSP\_004245, VSP\_004246;  
CC -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M60647; AAA42269.1; -;  
CC EMBL; J04035; AAA42268.1; -;  
CC EMBL; M86372; AAA42271.1; -;  
CC EMBL; M86355; AAA42271.1; JOINED.  
CC EMBL; M86363; AAA42271.1; JOINED.  
CC EMBL; M86364; AAA42271.1; JOINED.  
CC EMBL; M86366; AAA42271.1; JOINED.  
CC EMBL; M86371; AAA42271.1; JOINED.  
CC EMBL; M86376; AAA42272.1; -;  
CC EMBL; M86373; AAA42272.1; JOINED.  
CC EMBL; M86375; AAA42272.1; JOINED.  
CC PIR; A36106; EART.  
CC InterPro; IPR003979; tropoelastin.  
CC PRINTS; PR01500; TROPOELASTIN.  
CC Structural protein; Connective tissue; Repeat; Signal;  
KW Alternative splicing.  
FT NON TER 1 1  
FT SIGNAL <1 21 BY SIMILARITY.  
FT CHAIN 22 864 ELASTIN.  
FT DISULFID 854 859 BY SIMILARITY.  
FT VARSPLIC 263 307 Missing (in isoform 2, isoform 5, isoform  
7 and isoform 8).  
FT VARSPLIC 308 308 /FTID=VSP\_004244.  
FT VARSPLIC 809 823 Missing (in isoform 3, isoform 5, isoform  
6 and isoform 8).  
FT VARSPLIC 809 823 /FTID=VSP\_004245.  
FT VARSPLIC 809 823 Missing (in isoform 4, isoform 6, isoform  
7 and isoform 8).  
FT VARSPLIC 809 823 /FTID=VSP\_004246.  
SQ SEQUENCE 864 AA; 72786 MW; 456894BB09E79FD4 CRC64;  
  
Query Match 5.0%; Score 7; DB 1; Length 864;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 52 VLPGTGA 58  
DB 199 VLPGTGA 205

## RESULT 20

GCSP\_PEA STANDARD; PRT; 1057 AA.

AC P26969; 86

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Glycine dehydrogenase [decarboxylating], mitochondrial precursor (EC 1.4.4.2) (Glycine decarboxylase) (Glycine cleavage system P-protein)

DE GCSP OR GDCP.

GN Pisum sativum (Garden pea).

OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.

OX NCBI\_TaxID=3888; [1]

RN SEQUENCE FROM N.A.

RP STRAIN=cv. Birte; TISSUE=leaf;

RC MEDLINE=92184787; PubMed=1347530;

RA Turner S.R., Ireland R., Rawsthorne S.;

RT "Cloning and characterization of the P subunit of glycine decarboxylase from pea (Pisum sativum).";

RL J. Biol. Chem. 267:5355-5360(1992).

RN [2]

RP SEQUENCE OF 905-1057 FROM N.A.

RC STRAIN=cv. Alaska;

RA Shah K.S., Kim Y., Oliver D.J.;

RL Submitted (AUG-1990) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: The glycine cleavage system catalyzes the degradation of glycine. The P protein binds the alpha-amino group of glycine through its pyridoxal phosphate cofactor; CO(2) is released and the remaining methylene moiety is then transferred to the liposamide cofactor of the H protein.

CC -!- CATALYTIC ACTIVITY: Glycine + lipoylprotein = S-aminomethylidihydropyruvate + CO(2).

CC -!- COFACTOR: Pyridoxal phosphate.

CC -!- SUBUNIT: HOMODIMER. THE GLYCINE CLEAVAGE SYSTEM IS COMPOSED OF FOUR PROTEINS: P, T, L, AND H.

CC -!- SUBCELLULAR LOCATION: Mitochondrial.

CC -!- SIMILARITY: TO OTHER GLYCINE CLEAVAGE SYSTEM P-PROTEIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X59773; CAA42443.1; -

DR EMBL; X54377; CAA38252.1; -

DR PIR; A42109; A42109.

DR InterPro: IPR003437; GDC-P.

DR Pfam: PF02347; GDC-P; 1.

DR TIGRfams: TIGR00461; gcvp; 1.

KW Oxidoreductase; Mitochondrion; Pyridoxal phosphate; Transit peptide.

FT TRANSIT 1 86 MITOCHONDRION (POTENTIAL).

FT CHAIN 87 1057 GLYCINE DEHYDROGENASE (DECARBOXYLATING).

FT BINDING 792 792 PYRIDOXAL PHOSPHATE (BY SIMILARITY).

FT CONFLICT 905 906 I -> Y (IN REF. 2).

FT CONFLICT 919 919 P -> A (IN REF. 2).

SQ SEQUENCE 1057 AA; 114686 MW; 2FEA58E9A2AC47 CRC64;

Query Match 5.0%; Score 7; DB 1; Length 1057;

Best Local Similarity 100.0%; Pred. No. 47;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ESTSTTT 8

DB 27 ESTSTTT 33

## RESULT 21

RAD9\_YEAST STANDARD; PRT; 1309 AA.

AC P14737; Q04920;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE DNA repair protein Rad9.

GN RAD9 OR YDR217C OR YD9934.02C.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI\_TaxID=4932; [1]

RN SEQUENCE FROM N.A.

RP MEDLINE=89313732; PubMed=2664461;

RA Schiestl R.H., Reynolds P., Prakash S., Prakash L.;

RT "Cloning and sequence analysis of the Saccharomyces cerevisiae RAD9 gene and further evidence that its product is required for cell cycle arrest induced by DNA damage.";

RT Mol. Cell. Biol. 9:1882-1896(1989).

RL [2]

RN SEQUENCE FROM N.A., AND CHARACTERIZATION.

RP MEDLINE=91061763; PubMed=2247073;

RA Weinert T.A., Hartwell L.H.;

RT "Characterization of Rad9 of Saccharomyces cerevisiae and evidence that its function acts posttranslationally in cell cycle arrest after DNA damage.";

RL Mol. Cell. Biol. 10:6554-6564(1990).

RN [3]

RP SEQUENCE FROM N.A.

RA Murphy L., Harris D., Barrell B.G., Rajandream M.A.;

RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.

RN [4]

RP PHOSPHORYLATION, AND INTERACTION WITH RAD53.

RP MEDLINE=98429491; PubMed=9755168;

RA Vieland J.E., Gilbert C.S., Green C.M., Lowndes N.F.;

RT "The budding yeast Rad9 checkpoint protein is subjected to Mec1/Tel1-dependent hyperphosphorylation and interacts with Rad53 after DNA damage.";

RL EMBO J. 17:5679-5688(1998).

CC -!- FUNCTION: ESSENTIAL FOR CELL CYCLE ARREST AT THE G2 STAGE FOLLOWING DNA DAMAGE BY X-IRRADIATION OR INACTIVATION OF DNA LIGASE.

CC -!- SUBUNIT: Physically associates with RAD53.

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- SIMILARITY: Contains 1 BRCT domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M26049; AAA34954.1; -

DR EMBL; Z48612; CAA88497.1; -

DR PIR; S59444; BVBYD9.

DR PDB; 1FHR; 18-OCT-00.

DR PDB; 1J4K; 05-DEC-01.

DR PDB; 1J4L; 05-DEC-01.

DR PDB; 1J4P; 05-DEC-01.

DR PDB; 1J4Q; 05-DEC-01.

DR PDB; 1K2M; 05-DEC-01.

DR PDB; 1K2N; 05-DEC-01.

DR PDB; 1K3N; 05-DEC-01.

DR PDB; 1K3Q; 05-DEC-01.

DR SGD; S0002625; RAD9.

DR GO; GO:0005634; C:nucleus; IC.

DR GO; GO:0005515; P:protein binding activity; IDA.

DR GO; GO:0000077; P:DNA damage checkpoint; IMP.

```
DR GO; GO:0006289; P:nucleotide-excision repair; IMP.
DR GO; GO:0000074; P:regulation of cell cycle; IGI.
DR InterPro; IPR001357; BRCT.
DR Pfam; PF00533; BRCT; 1.
DR SMART; SM00292; BRCT; 1.
DR PROSITE; PS00172; BRCT; 1.
KW Cell cycle; DNA damage; DNA replication inhibitor; Nuclear protein;
KW Phosphorylation; 3D-structure.
FT DOMAIN 994 1122
FT CONFLICT 433 433 C -> S (IN REF. 3).
SQ SEQUENCE 1309 AA; 148413 MW; 6B77D39A95021F84 CRC64;

Query Match 5.0%; Score 7; DB 1; Length 1309;
Best Local Similarity 100.0%; Pred.No.58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 RHLRSLK 133
DB 1082 RHLRSLK 1088

RESULT 22
RPOP_HALN1 STANDARD; PRT; 44 AA.
AC Q9HSG9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE RPOP OR VNG0237H. (strain NRC-1 / ATCC 700922 / JCM 11081).
OS Halobacterium sp.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng M.V., Kennedy S.P., Maniatis G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Shrogha J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Ledhauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Beck R.F., Pohlshocher M., Spudis J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
CC of DNA into RNA using the four ribonucleoside triphosphates as
CC substrates.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA)(N)
CC -!- SIMILARITY: BELONGS TO THE ARCHAEABACTERIA RPOP / EUKARYOTIC RPOP10
CC RNA POLYMERASE SUBUNIT FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AE004988; AAG18837.1; .
DR PIR; A84184; A84184.
DR HAMAP; MF 00615; -; 1.
DR InterPro; IPR006591; RNA_pol_Rbp10.
DR SMART; SM00659; RPOP1C; 1.
KW Transferase; DNA-directed RNA polymerase; Transcription;
KW Metal-binding; Zinc-finger; Complete proteome.
FT ZN FING 5 25 C4-TYPE (POTENTIAL).
SQ SEQUENCE 44 AA; 5179 MW; C175E83AA4CADFAC CRC64;

GO; GO:0006289; P:nucleotide-excision repair; IMP.
GO; GO:0000074; P:regulation of cell cycle; IGI.
InterPro; IPR001357; BRCT.
Pfam; PF00533; BRCT; 1.
SMART; SM00292; BRCT; 1.
PROSITE; PS00172; BRCT; 1.
Cell cycle; DNA damage; DNA replication inhibitor; Nuclear protein;
Phosphorylation; 3D-structure.
DOMAIN 994 1122
CONFLICT 433 433 C -> S (IN REF. 3).
SEQUENCE 1309 AA; 148413 MW; 6B77D39A95021F84 CRC64;

Query Match 5.0%; Score 7; DB 1; Length 1309;
Best Local Similarity 100.0%; Pred.No.58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 RHLRSLK 133
DB 1082 RHLRSLK 1088

RESULT 22
RPOP_HALN1 STANDARD; PRT; 44 AA.
AC Q9HSG9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE RPOP OR VNG0237H. (strain NRC-1 / ATCC 700922 / JCM 11081).
OS Halobacterium sp.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng M.V., Kennedy S.P., Maniatis G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Shrogha J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Ledhauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Beck R.F., Pohlshocher M., Spudis J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
CC of DNA into RNA using the four ribonucleoside triphosphates as
CC substrates.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA)(N)
CC -!- SIMILARITY: BELONGS TO THE ARCHAEABACTERIA RPOP / EUKARYOTIC RPOP10
CC RNA POLYMERASE SUBUNIT FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AE004988; AAG18837.1; .
DR PIR; A84184; A84184.
DR HAMAP; MF 00615; -; 1.
DR InterPro; IPR006591; RNA_pol_Rbp10.
DR SMART; SM00659; RPOP1C; 1.
KW Transferase; DNA-directed RNA polymerase; Transcription;
KW Metal-binding; Zinc-finger; Complete proteome.
FT ZN FING 5 25 C4-TYPE (POTENTIAL).
SQ SEQUENCE 44 AA; 5179 MW; C175E83AA4CADFAC CRC64;

Query Match 4.3%; Score 6; DB 1; Length 44;
Best Local Similarity 100.0%; Pred.No.34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 CGHRVL 53
DB 25 CGHRVL 30

RESULT 23
ATP8_ASTPE STANDARD; PRT; 54 AA.
AC Q3882;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP synthase protein 8 (EC 3.6.3.14) (ATPase subunit 8) (A6L).
GN MTATP8 OR ATP8
OS Asterina pectinifera (Starfish).
OC Mitochondrion.
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Asterozoa;
OC Asteroidea; Valvatacea; Valvatida; Asterinidae; Asterina.
OX NCBI_TaxID=7594;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95402698; PubMed=7672576;
RA Asakawa S., Himeno H., Miura K.-I., Watanabe K.;
RA "Nucleotide sequence and gene organization of the starfish Asterina
RA pectinifera mitochondrial genome.";
RL Genetics 140:1047-1060(1995).
CC -!- FUNCTION: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT
CC (CF(0) SUBUNIT) OF THE MITOCHONDRIAL ATPASE COMPLEX.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(in) = ADP + phosphate +
CC H(+)(out).
CC -!- SUBCELLULAR LOCATION: Membrane-bound.
CC -!- SIMILARITY: BELONGS TO THE ATPASE PROTEIN 8 FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; D16387; BA03883.1; .
DR PIR; S70600; S70600.
DR InterPro; IPR001421; ATPase8_mit.
DR Pfam; PF00895; ATP-synt_8; 1.
DR KW Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
SQ SEQUENCE 54 AA; 6241 MW; 9EABDACEB9CDF5F1 CRC64;

Query Match 4.3%; Score 6; DB 1; Length 54;
Best Local Similarity 100.0%; Pred.No.34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TSTTTN 9
DB 39 TSTTTN 44

RESULT 24
RKL4_OENAM STANDARD; PRT; 74 AA.
ID RKL4_OENAM
AC P42340;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Chloroplast 50S ribosomal protein L14 (Fragments).
GN RPL14.
OS Oenothera amplexifolia (Evening primrose).
```

DR	SMART; SW00389; HOK; 1.
DR	PROSITE; PS0027; HOMEBOX_1; 1.
DR	PROSITE; PSS0071; HOMEBOX_2; 1.
KW	Homeobox; DNA-binding; Nuclear protein; Multigene family.
FT	NON_TER 1 1
FT	DOMAIN 2 25 ELK DOMAIN.
FT	DNA_BIND 26 88 HOMEBOX (TALE-TYPE).
FT	DOMAIN 27 32 POLY-LYS.
FT	NON_TER 88 88
SQ	SEQUENCE 88 AA; 10789 MW; CEBFEDD754A024C2 CRC64;
 Query Match 4.3%; Score 6; DB 1; Length 89; Best Local Similarity 100.0%; Pred.No. 54; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	88 RELSRK 93 
Db	22 RELSRK 27 
 RESULT 26	
MMOD	METTR STANDARD; PRT; 102 AA.
ID	_NMOT METTR STANDARD; PRT; 102 AA.
AC	OS352; 2003 (Rel. 41, Created)
DT	28-FEB-2003 (Rel. 41, Last sequence update)
DT	28-FEB-2003 (Rel. 41, Last annotation update)
DE	Methane monooxygenase component D.
GN	MMOD.
OS	Methylosinus trichosporium.
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC	Methylcytaceae; Methylosinus.
OX	NCBI_TaxId=426;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=OB3b;
RX	MEDLINE=92153031; PubMed=1785954;
RA	Cady D.L.N., Laidler V., Salmond G.P.C., Murrell J.C.;
RT	"The methane monooxygenase gene cluster of Methylosinus trichosporium:
RT	cloning and sequencing of the mmoC gene.";
EL	Arch. Microbiol. 156:477-483(1991).
CC	-1- FUNCTION: Not known.
CC	-1- SUBUNIT: The soluble methane monooxygenase (sMMO) consists of four
CC	components A/MMOH (composed of alpha/mmoX, beta/mmoY and
CC	gamma/mmoZ), B/MMOB (mmoB), C/MMOR (mmoC) and D/MMOD (mmoD).
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to licenses@isb-sib.ch).
CC	-----
CC	EMBL; S81887; AAB21392.1; --
DR	PJR; B48360; B48360.
KW	Oxidoreductase; Monooxygenase.
SQ	SEQUENCE 102 AA; 11926 MW; 055307622A09409D CRC64;
 Query Match 4.3%; Score 6; DB 1; Length 102; Best Local Similarity 100.0%; Pred.No. 62; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	102 SQESAR 107 
Db	49 SQESAR 54 
 RESULT 27	
ACCD	ORYSA
ID	_ACCD ORYSA STANDARD; PRT; 106 AA.
PI	P1221g;
OT	01-OCT-1988 (Rel. 12, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta  
DE (EC 6.4.1.2) (ACCASE beta chain).  
GN ACCD OR YCF11.  
OS Oryza sativa (Rice).  
OS Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OX NCBI\_TaxID=4530;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RX MEDLINE=89364698; PubMed=2770692;  
RA Hiratsuka J., Shimada H., Whittier R., Ishibashi T., Sakamoto M.,  
RA Mori M., Kondo C., Honji Y., Sun C.-R., Meng B.-Y., Li Y.-Q.,  
RA Kanno A., Nishikawa Y., Hirai A., Shinozaki K., Sugita M.,  
RT "The complete sequence of the rice (*Oryza sativa*) chloroplast genome:  
RT intermolecular recombination between distinct trna genes accounts for  
RT a major plastid DNA inversion during the evolution of the cereals.";  
RL Mol. Gen. Genet. 217:185-194(1989).  
RL -!- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3)(-) = ADP + phosphate  
CC + malonyl-CoA.  
CC -!- SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.  
CC -!- CAUTION: CORRESPONDS TO THE C-TERMINAL PART OF OTHER PLANTS  
CC ACCD.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X15901; CAA33956.1; -.  
DR PIR; J00234; J00234.  
DR Granene; P12218; -.  
KW Fatty acid biosynthesis; Ligase; Chloroplast.  
SQ SEQUENCE 106 AA; 12455 MW; D0410243163BEP2E CRC64;  
-----  
Query Match 4.3%; Score 6; DB 1; Length 106;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
-----  
QY 80 QSTKSP 85  
Db 64 QSTKSP 69  
-----  
RESULT 28  
SSI\_STRGI STANDARD; PRT; 107 AA.  
AC P28592;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DE 01-JUN-1994 (Rel. 29, Last annotation update)  
DE Alkaline protease inhibitor 2C' (API-2C').  
OS Streptomyces griseolincarnatus.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomycetes.  
OX NCBI\_TaxID=29305;  
RN (1)  
RP SEQUENCE.  
RA Suzuki K., Uyeda M., Shibata M.;  
RT "Partial amino acid sequence of an alkaline protease inhibitor, API-2  
RT (b and c).";  
RL Agric. Biol. Chem. 45:629-634(1981).  
CC -!- FUNCTION: STRONG INHIBITORY ACTIVITY TOWARD SUBTILISIN.  
CC -!- SUBUNIT: Homodimer.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- SIMILARITY: BELONGS TO THE SERINE PROTEASE INHIBITORS SSI FAMILY.

DR PIR; PC1261; PC1261.  
DR HSP; P01006; 2SIC.  
DR InterPro; IPR000691; Strep\_subt\_inhib.  
DR Pfam; PF00720; SSI; 1.  
DR PRINTS; PR00234; SSBTLINHSTR.  
DR ProDom; PD004028; Strep\_subt\_inhib; 1.  
DR PROSITE; PS00999; SSI; 1.  
KW Serine protease inhibitor.  
FT DISULFID 29 44 BY SIMILARITY.  
FT DISULFID 65 95 BY SIMILARITY.  
FT ACT SITE 67 68 REACTIVE BOND.  
SQ SEQUENCE 107 AA; 10972 MW; 76043BA9F876D1CA CRC64;  
-----  
Query Match 4.3%; Score 6; DB 1; Length 107;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
-----  
QY 62 AATVTP 67  
Db 16 AATVTP 21  
-----  
RESULT 29  
YHEA\_RHOCA STANDARD; PRT; 124 AA.  
AC P29962;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DE Hypothetical 12.7 kDa protein in hella 5' region (ORF124).  
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;  
OC Rhodobacteraceae; Rhodobacter.  
OX NCBI\_TaxID=1061;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=SB1003 / St Louis;  
RX MEDLINE=92146961; PubMed=1310666;  
RA Beckman D.L., Trawick D.R., Kranz R.G.;  
RL "Bacterial cytochromes c biogenesis.";  
RL Genes Dev. 6:268-283(1992).  
CC -!- FUNCTION: NOT REQUIRED FOR THE BIOGENESIS OF C-TYPE CYTOCHROMES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X63462; CAA45060.1; -.  
DR PIR; S23662; S23662.  
DR InterPro; IPR006729; DUF598.  
DR Pfam; PF04635; DUF598; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 124 AA; 12711 MW; 1CA4FDFD6F77DFCE CRC64;  
-----  
Query Match 4.3%; Score 6; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
-----  
QY 41 LRALAG 46  
Db 52 LRALAG 57  
-----  
RESULT 30  
ABAL\_TRIAB STANDARD; PRT; 131 AA.  
ID ABAL\_TRIAB  
AC P8111;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Alboaggregin A subunit 1.  
OS Trimeresurus albolabris (White-lipped pit viper).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Lepidosaurs; Squamata; Scleroglossa; Serpentes; Colubroidea;  
OC Viperidae; Crotalinae; Trimeresurus.  
OX NCBI\_TaxID=9765;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Venom;  
RX MEDLINE=98189535; PubMed=9531050;  
RA Kowalek M.A., Tan L., Holt J.C., Peng M., Karczewski J.,  
RA Galvete J.J., Niewiarowski S.;  
RT "Alboaggregins A and B. Structure and interaction with human  
RT platelets";  
RL Thromb. Haemost. 79:609-613(1998).  
CC -1- FUNCTION: Binds to platelet GPIb/IX receptor system and stimulates  
CC agglutination.  
CC -1- SUBUNIT: Heterotetramer of the subunits 1, 2, 3 and 4,  
CC disulfide-linked.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.  
DR HSSP; P23806; 11XX.  
DR InterPro; IPR001304; Lectin\_C.  
DR Pfam; PF00059; lectin\_C; 1.  
DR SMART; SM00034; CLECT; 1.  
DR PROSITE; PS00615; C\_TYPE\_LLECTIN\_1; 1.  
DR PROSITE; PS50041; C\_TYPE\_LLECTIN\_2; 1.  
DR LECTIN.  
KW DOMAIN.  
FT DOMAIN 1 129 C-TYPE LECTIN (LONG FORM).  
FT DISULFID 2 13 BY SIMILARITY.  
FT DISULFID 30 127 BY SIMILARITY.  
FT DISULFID 102 119 BY SIMILARITY.  
SQ SEQUENCE 131 AA; 15427 MW; B3569F5BF91F6624 CRC64;  
Query Match 4.3%; Score 6; DB 1; Length 131;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 119 LEKRAE 124  
Db 105 LEKRAE 110  
RESULT 31  
RL24\_AERPE STANDARD; PRT; 132 AA.  
AC QYF83;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 18-OCT-2001 (Rel. 40, Last annotation update)  
DE 50S ribosomal protein L24p.  
GN RPL24P OR APE0358.  
OS Aeropyrum pernix.  
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;  
OC Desulfurococaceae; Aeropyrum.  
OX NCBI\_TaxID=56636;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K1;  
RX MEDLINE=99310339; PubMed=10382966;  
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,  
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,  
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,  
RA Yamazaki J., Kushima N., Oguchi A., Aoki K.-I., Kubota K.,  
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;  
RT "Complete genome sequence of an aerobic hyper-thermophilic  
RT crenarchaeon, Aeropyrum pernix K1";  
RL DNA Res. 6:83-101(1999).  
CC -1- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AP000059; BAA79313.1; -  
DR PIR; E72727; E72727.  
DR InterPro; IPR005824; KOW.  
DR InterPro; IPR006646; KOW sub.  
DR InterPro; IPR005756; L26\_euk\_arch.  
DR InterPro; IPR005825; Ribosomal\_L24\_26.  
DR Pfam; PF00467; KOW; 1.  
DR SMART; SM00739; KOW; 1.  
DR TIGRFAMs; TIGR01080; rplX A E; 1.  
DR PROSITE; PS01108; RIBOSOMAL\_L24; 1.  
KW Ribosomal protein; Complete\_protosome.  
SQ SEQUENCE 132 AA; 15129 MW; 58A83CE320F3FC96 CRC64;  
Query Match 4.3%; Score 6; DB 1; Length 132;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 84 SPELRE 89  
Db 34 SPELRE 39  
RESULT 32  
APB\_RABIT STANDARD; PRT; 144 AA.  
ID APB\_RABIT  
AC P17165;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-AUG-1990 (Rel. 15, Last annotation update)  
DE Apolipoprotein B (Fragment).  
GN APOB.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88335569; PubMed=3419914;  
RA Sudarickov A., Surguchov A.;  
RT "Probe for rabbit apolipoprotein B gene.";  
RL Nucleic Acids Res. 16:8187-8187(1988).  
CC -1- FUNCTION: APOLIPOPROTEIN B IS A MAJOR PROTEIN CONSTITUENT OF  
CC CHYLOMICRONS, VLDL AND LDL. IT FUNCTIONS AS A RECOGNITION  
CC SIGNAL FOR THE CELLULAR BINDING AND INTERNALIZATION OF LDL  
CC PARTICLES BY THE APOB/E RECEPTOR.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X07480; CAA30366.1; -  
DR Plasma; Lipid transport; VLDL; LDL; Chylomicron; Heparin-binding;  
KW Cholesterol metabolism; Glycoprotein; Atherosclerosis.  
FT NON TER 1 1  
FT NON TER 144 144  
SQ SEQUENCE 144 AA; 15664 MW; 5ED9F09D0A9EFA26 CRC64;  
Query Match 4.3%; Score 6; DB 1; Length 144;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 69 GASMKL 74



```

Db          92 GASKML 97
|||||
RESULT 33
VP8_BPPH6   STANDARD;          PRT; 149 AA.
AC P07579;
DT 01-APR-1998 (Rel. 07, Created)
DT 01-APR-1998 (Rel. 07, Last sequence update)
DT 01-JAN-1990 (Rel. 13, Last annotation update)
DE Major nucleocapsid protein (P8 protein).
GN P8.
OS Bacteriophage phi-6.
OC Viruses; dsRNA viruses; Cystoviridae; Cystovirus.
OX NCBI_TaxID=10879;
RN [1]
RX MEDLINE=86144095; PubMed=3754015;
RA McGraw T., Mindich L., Frangione B.;
RT "Nucleotide sequence of the small double-stranded RNA segment of
RT bacteriophage phi 6: novel mechanism of natural translational
RT control."
RL J. Virol. 58:142-151(1986).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M12921; AAA32358.1; -.
DR F01; A23368; VHBPF6.
KW Nucleocapsid.
SQ SEQUENCE 149 AA; 16004 MW; ECED0D195186A929 CRC64;

Query Match 4.3%; Score 6; DB 1; Length 149;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 SAIATAT 64
   |||||
Db 16 SAIATAT 21

RESULT 34
HES2_MOUSE
ID HES2_MOUSE STANDARD;          PRT; 157 AA.
AC O54792;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transcription factor HES-2 (Hairy and enhancer of split 2).
GN HES2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/J;
RX MEDLINE=98234545; PubMed=95703950;
RA Nishimura M., Isaka F., Ishibashi M., Tomita K., Tsuda H.,
RA Nakanishi S., Kageyama R.;
RT "Structure, chromosomal locus, and promoter of mouse HES2 gene, a
RT homologue of Drosophila hairy and Enhancer of split."
RL Genomics 49:69-75(1998).
CC -----
CC -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR OF GENES THAT REQUIRE A BHLH
CC PROTEIN FOR THEIR TRANSCRIPTION.
CC -!- SUBUNIT: TRANSCRIPTION REPRESSION REQUIRES FORMATION OF A COMPLEX
CC WITH A CO-REPRESSOR PROTEIN (GROUCHO).

```

```

CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DOMAIN: HAS A PARTICULAR TYPE OF BASIC DOMAIN (PRESENCE OF A
CC HELIX-INTERRUPTING PROLINE) THAT BINDS TO THE N-BOX (CACNAG),
CC RATHER THAN THE CANONICAL E-BOX (CANNIG).
CC -!- DOMAIN: THE CARBOXYL-TERMINAL WRW MOTIF IS A TRANSCRIPTIONAL
CC REPRESSION DOMAIN NECESSARY FOR THE INTERACTION WITH GROUCHO. A
CC TRANSCRIPTIONAL CO-REPRESSOR RECRUITED TO SPECIFIC TARGET DNA BY
CC HAIRY-RELATED PROTEINS.
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS "HAIRY" SUBFAMILY.
CC -!- SIMILARITY: Contains 1 orange domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB009967; BAA24091.1; -.
DR MGD; MGI:1098624; Hes2.
DR InterPro; IPR001092; HLH basic.
DR InterPro; IPR003650; Orange.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR SMART; SM00511; ORANGE; 1.
DR PROSITE; PS00038; HLH 1; 1.
DR PROSITE; PS00888; HLH 2; 1.
KW Transcription regulation; DNA-binding; Nuclear protein; Repressor.
FT DNA_BIND 13 26
   BASIC DOMAIN.
FT DOMAIN 27 71
   HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 139 148
   PRO-RICH.
FT DOMAIN 154 157
   WRPW MOTIF (REQUIRED FOR ACTIVITY)
   (BY SIMILARITY).
SQ SEQUENCE 157 AA; 17231 MW; 570A0C67F4992EA7 CRC64;

Query Match 4.3%; Score 6; DB 1; Length 157;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 PLEKXR 122
   |||||
Db 19 PLEKXR 24

RESULT 35
HES2_RAT
ID HES2_RAT STANDARD;          PRT; 157 AA.
AC F35429;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transcription factor HES-2 (Hairy and enhancer of split 2).
GN HES2 OR HES-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryonic brain;
RX MEDLINE=93358866; PubMed=8354270;
RA Ishibashi M., Sasai Y., Nakanishi S., Kageyama R.;
RT "Molecular characterization of HES-2, a mammalian helix-loop-helix
RT factor structurally related to Drosophila hairy and Enhancer of
RT split."
RL Eur. J. Biochem. 215:645-652(1993).
CC -----
CC -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR OF GENES THAT REQUIRE A BHLH
CC PROTEIN FOR THEIR TRANSCRIPTION.
CC -!- SUBUNIT: TRANSCRIPTION REPRESSION REQUIRES FORMATION OF A COMPLEX
CC WITH A CO-REPRESSOR PROTEIN (GROUCHO).
CC -!- SUBCELLULAR LOCATION: Nuclear.

```

CC -!- DOMAIN: HAS A PARTICULAR TYPE OF BASIC DOMAIN (PRESENCE OF A  
 CC HELIX-INTERUPTING PROLINE) THAT BINDS TO THE N-BOX (CACNAG),  
 CC RATHER THAN THE CANONICAL E-BOX (CANNTG).  
 CC -!- DOMAIN: THE CARBOXYL-TERMINAL WRPW MOTIF IS A TRANSCRIPTIONAL  
 CC REPRESSION DOMAIN NECESSARY FOR THE INTERACTION WITH GROUCHO. A  
 CC TRANSCRIPTIONAL CO-REPRESSOR RECRUITED TO SPECIFIC TARGET DNA BY  
 CC HAIRY-RELATED PROTEINS.  
 CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS. "HAIRY" SUBFAMILY.  
 CC -!- SIMILARITY: Contains 1 orange domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D14029; BAA03118.1; -;  
 CC PIR; S35037; S35037.  
 CC TRANSPAC; T01650; -;  
 CC InterPro; IPR001092; HLH basic.  
 CC InterPro; IPR003650; Orange.  
 CC Pfam; PF00010; HLH; 1.  
 CC SMART; SM00353; HLH; 1.  
 CC SMART; SM00511; ORANGE; 1.  
 CC PROSITE; PS00038; HLH 1; 1.  
 CC PROSITE; PS00885; HLH 2; 1.  
 CC Transcription regulation; DNA-binding; Nuclear protein; Repressor.  
 CC DNA BIND 13 26  
 CC BASIC DOMAIN.  
 CC HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 CC PRO-RICH.  
 CC WRPW MOTIF (REQUIRED FOR ACTIVITY)  
 CC (BY SIMILARITY).  
 CC SEQUENCE 157 AA; 17028 MW; B5D621E814AE0369 CRC64;  
 CC  
 CC Query Match 4.3%; Score 6; DB 1; Length 157;  
 CC Best Local Similarity 100.0%; Pred. No. 92;  
 CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 117 PLEKR 122  
 CC |||||  
 CC 19 PLEKR 24  
 CC  
 CC RESULT 36  
 CC YRNS\_CAEEL STANDARD; PRT; 160 AA.  
 CC AC Q09419;  
 CC DT 01-NOV-1995 (Rel. 32, Created)  
 CC DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 CC DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 CC DE Hypothetical 16.4 kDa protein R07B1.5 in chromosome X precursor.  
 CC GN R07B1.5.  
 CC OS Caenorhabditis elegans.  
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 CC Rhabditidae; Peloderinae; Caenorhabditis.  
 CC NCBI\_TaxID=6239;  
 CC [1]  
 CC RN SEQUENCE FROM N.A.  
 CC RC STRAIN=Bristol N2;  
 CC Rershaw J.;  
 CC RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: SOME, TO C.ELEGANS C30G12.4.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

CC EMBL; Z48621; CAA89542.1; -;  
 CC PIR; T23995; T23995.  
 CC WormPep; R07B1.5; CE01631.  
 CC InterPro; IPR002601; C6.  
 CC Pfam; PF01681; C6; 1.  
 CC Hypothetical protein; Signal.  
 CC FT SIGNAL 1 18  
 CC CHAIN 19 160  
 CC SER/THR-RICH.  
 CC FT DOMAIN 22 58  
 CC SEQUENCE 160 AA; 16413 MW; E9C21E219712A13 CRC64;  
 CC  
 CC Query Match 4.3%; Score 6; DB 1; Length 160;  
 CC Best Local Similarity 100.0%; Pred. No. 93;  
 CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 3 STSTTT 8  
 CC |||||  
 CC 44 STSTTT 49  
 CC  
 CC RESULT 37  
 CC GTR4\_CANFA  
 CC ID GTR4\_CANFA STANDARD; PRT; 162 AA.  
 CC AC Q9XST2;  
 CC DT 28-FEB-2003 (Rel. 41, Created)  
 CC DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DE Solute carrier family 2, facilitated glucose transporter, member 4  
 CC (Glucose transporter type 4, insulin-responsive) (Fragment).  
 CC GN SLC2A4 OR GLUT4.  
 CC OS Canis familiaris (Dog).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 CC NCBI\_TaxID=9615;  
 CC RN [1]  
 CC SEQUENCE FROM N.A.  
 CC TISSUE=Thyroid;  
 CC MEDLINE=20422104; PubMed=10964405;  
 CC Pichon B., Mercan D., Pouillon V., Christophe-Hobertus C.,  
 CC Christophe D.;  
 CC RT "A method for the large-scale cloning of nuclear proteins and nuclear  
 CC targeting sequences on a functional basis.";  
 CC RL Anal. Biochem. 284:231-239(2000).  
 CC -!- FUNCTION: Insulin-regulated facilitative glucose transporter.  
 CC -!- SUBUNIT: Binds to DAXX (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Localizes  
 CC primarily to the perinuclear region, undergoing continued  
 CC recycling to the plasma membrane where it is rapidly  
 CC reinternalized. The dileucine internalization motif is critical  
 CC for intracellular sequestration (By similarity).  
 CC -!- PTM: Sumoylated (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. GLUCOSE  
 CC TRANSPORTERS SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AJ389533; CAB46835.1; -;  
 CC InterPro; IPR005828; Sub transporter.  
 CC InterPro; IPR005829; Sug transporter.  
 CC Pfam; PF00083; sugar tr\_1  
 CC PRINTS; PR00171; SUGTRANSPORT.  
 CC PROSITE; PS00216; SUGAR\_TRANSPORT\_1; 1.  
 CC PROSITE; PS00217; SUGAR\_TRANSPORT\_2; PARTIAL.  
 CC Transmembrane; Sugar Transport; Transport; Multigene family.  
 CC NON\_TER 1 1  
 CC FT

FT DOMAIN <1 13 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 14 34 8 (POTENTIAL).  
 FT DOMAIN 35 43 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 44 64 9 (POTENTIAL).  
 FT DOMAIN 65 75 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 76 96 10 (POTENTIAL).  
 FT DOMAIN 97 107 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 108 128 11 (POTENTIAL).  
 FT DOMAIN 129 135 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 136 156 12 (POTENTIAL).  
 FT DOMAIN 157 >162 CYTOPLASMIC (POTENTIAL).  
 FT NON TER 162 162  
 SQ SEQUENCE 162 AA; 17453 MW; 0C58CBB23C6AD2BB CRC64;

Query Match 4.3%; Score 6; DB 1; Length 162;  
 Best Local Similarity 100.0%; Pred. No. 94;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 44 LAGMCG 49  
 Db 47 LAGMCG 52

## RESULT 38

B3AR MERUN STANDARD; PRT; 167 AA.  
 ID B3AR MERUN  
 AC 070432;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Beta-3 adrenergic receptor (Fragment).  
 GN ADRB3.  
 OS Meriones unguiculatus (Mongolian jird) (Mongolian gerbil).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;  
 OC Meriones.  
 CC NCBI\_TaxID=10047;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain, Fat, and Vestibular labyrinth;  
 RX MEDLINE=99329129; PubMed=10398761;  
 RA Wangemann P., Liu J., Shinozono M., Scofield M.A.;  
 RT "Beta1-adrenergic receptors but not beta2-adrenergic or vasopressin  
 RT receptors regulate K+ secretion in vestibular dark cells of the inner  
 RT ear.";  
 RL J. Membr. Biol. 170:67-77(1999).  
 CC -!- FUNCTION: BETA-ADRENERGIC RECEPTORS MEDIATE THE CATECHOLAMINE-  
 CC INDUCED ACTIVATION OF ADENYLATE CYCLASE THROUGH THE ACTION OF G  
 CC PROTEINS. BETA-3 IS INVOLVED IN THE REGULATION OF LIPOLYSIS AND  
 CC THERMOGENESIS.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AF055351; AAC12769.1;  
 DR InterPro; IPR000276; GPCR\_Rhodpsn.  
 DR Pfam; PF00001; 7tm1.1;  
 DR PRINTS; PS00237; GPCRHOOPS  
 DR PROSITE; PS00237; G-PROTEIN\_RECEP\_F1\_1; PARTIAL.  
 DR PROSITE; PS00262; G-PROTEIN\_RECEP\_F1\_2; 1.  
 KW G-protein coupled receptor; Transmembrane; Multigene family.  
 FT NON TER 1 1  
 FT DOMAIN <1 25 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 26 47 5 (POTENTIAL).  
 FT DOMAIN 48 114 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 115 136 6 (POTENTIAL).

FT DOMAIN 137 148 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 149 >167 7 (POTENTIAL).  
 FT NON TER 167 167  
 SQ SEQUENCE 167 AA; 18324 MW; CEA78CF68DEEB1C1 CRC64;

Query Match 4.3%; Score 6; DB 1; Length 167;  
 Best Local Similarity 100.0%; Pred. No. 97;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 LRALAG 46  
 Db 136 LRALAG 141

## RESULT 39

ILVH MYCAV STANDARD; PRT; 167 AA.  
 ID ILVH MYCAV  
 AC Q59439;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Acetolactate synthase small subunit (EC 2.2.1.6) (AHAS)  
 DE (Acetohydroxy-acid synthase small subunit) (ALS).  
 GN ILVH OR ILVN.  
 OS Mycobacterium avium.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 CC NCBI\_TaxID=1764;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97080504; PubMed=8921849;  
 RA Gusberti L., Cantoni R., de Rossi E., Branzoni M., Riccardi G.;  
 RT "Cloning and sequencing of the ilvNC gene cluster from Mycobacterium  
 RT avium.";  
 RL Gene 177:83-85 (1996).  
 CC -!- CATALYTIC ACTIVITY: 2 pyruvate = 2-acetolactate + CO(2).  
 CC -!- PATHWAY: Valine and isoleucine biosynthesis; first step.  
 CC -!- SUBUNIT: DIMER OF LARGE AND SMALL CHAINS (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE ACETOLACTATE SYNTHASE SMALL SUBUNIT  
 CC FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; L49392; AAB38427.1;  
 DR InterPro; IPR004789; Acolac\_sm.  
 DR InterPro; IPR002912; ACT.  
 DR Pfam; PF01842; ACT; 1.  
 DR ProDom; PD002844; Acolac\_sm; 1.  
 DR TIGRFAMs; TIGR00119; acolac\_sm; 1.  
 KW Transferase; Branched-chain amino acid biosynthesis.  
 SQ SEQUENCE 167 AA; 18131 MW; 2F234C6CE9B8DD2D CRC64;

Query Match 4.3%; Score 6; DB 1; Length 167;  
 Best Local Similarity 100.0%; Pred. No. 97;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 EALLRV 32  
 Db 135 EALLRV 140

## RESULT 40

ILVH MYCTU STANDARD; PRT; 168 AA.  
 ID ILVH MYCTU  
 AC O53219;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Acetolactate synthase small subunit (EC 2.2.1.6) (AHAS)  
DE (Acetoxyhydroxy-acid synthase small subunit) (ALS).  
GN ILVH OR RV3002C OR MT3082 OR MTV012.16C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -|- CATALYTIC ACTIVITY: 2 pyruvate = 2-acetolactate + CO(2).  
CC -|- PATHWAY: Valine and isoleucine biosynthesis; first step.  
CC -|- SUBUNIT: DIMER OF LARGE AND SMALL CHAINS (BY SIMILARITY).  
CC -|- SIMILARITY: BELONGS TO THE ACETOLACTATE SYNTHASE SMALL SUBUNIT  
CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AL021287; CAAl6087.1; -;  
DR EMBL; AE007128; AAK47411.1; -;  
DR PIR; E70855; E70855.  
DR TIGR; MT3082; -;  
DR TuberculList; RV3002c; -;  
DR InterPro; IPR004789; Acolac\_sm.  
DR InterPro; IPR002912; ACT.  
DR Pfam; PF01842; ACT; 1.  
DR ProDom; PD002844; Acolac\_sm; 1.  
DR TIGRFAMS; TIGR00119; acolac\_sm; 1.  
KW Transferase; Branched-chain amino acid biosynthesis;  
KW Complete proteome.  
SQ SEQUENCE 168 AA; 18187 MW; 523488D6114AB354 CRC64;

Query Match 4.3%; Score 6; DB 1; Length 168;  
Best Local Similarity 100.0%; Pred. No. 98;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 EALLRV 32  
| | | | |  
Db 135 EALLRV 140

Search completed: November 14, 2003, 10:51:04  
Job time : 30 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 10:48:21 ; Search time 34 Seconds

(without alignments)  
1070.159 Million cell updates/sec

Title: US-10-087-573-2

Perfect score: 141

Sequence: 1 MESTSTTNFVAENRPTTGE.....RAEYFRHLRLSKSQGVNRLI 141

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

- SPTREMBL.23.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_prodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	141	100.0	141	5 Q8MNM4	Q8mnm4 babesia can
2	119	84.4	285	5 Q8MQH4	Q8mqh4 babesia can
3	118	84.4	285	5 Q8MQH4	Q8mqh4 babesia can
4	8	5.7	334	10 Q9LHS1	Q9lhs1 arabidopsis
5	8	5.7	434	10 Q9FG07	Q9fg07 arabidopsis
6	8	5.7	513	10 Q9STM8	Q9stm8 arabidopsis
7	8	5.7	534	16 Q8EGU2	Q8ecu2 shewanella
8	8	5.7	559	10 Q9SK35	Q9sk35 arabidopsis
9	8	5.7	609	10 Q9RH26	Q9rh26 arabidopsis
10	8	5.7	689	10 Q9FG06	Q9fg06 arabidopsis
11	7	5.0	707	10 Q9STM7	Q9stm7 arabidopsis
12	7	5.0	65	16 Q9RDQ2	Q9rdq2 streptomyces
13	7	5.0	91	5 Q9BNJ8	Q9bnj8 platydesmus
14	7	5.0	105	2 Q49812	Q49812 mycobacteri
15	7	5.0	133	16 Q8DF92	Q8df92 vibrio vuln
16	7	5.0	148	2 Q49876	Q49876 mycobacteri
					Q9vct2 drosophila

17	7	5.0	172	6 Q9BDZ0	Q9bdz0 macaca mula
18	7	5.0	203	5 Q9VH39	Q9vnh39 drosophila
19	7	5.0	223	6 Q9SLB5	Q9slb5 ovis aries
20	7	5.0	225	16 Q8FWQ6	Q8fwq6 bruceella su
21	7	5.0	237	16 Q8YBM1	Q8ybm1 bruceella me
22	7	5.0	243	16 Q987C4	Q987c4 rhizobium l
23	7	5.0	255	5 Q95SR4	Q95sr4 drosophila
24	7	5.0	269	5 Q93UL3	Q93ul3 rhizobium g
25	7	5.0	270	2 Q93UK8	Q93uk8 sinorhizobi
26	7	5.0	273	2 Q8RNL4	Q8rnl4 sinorhizobi
27	7	5.0	281	2 Q93UL5	Q93ul5 rhizobium e
28	7	5.0	282	2 Q93UL4	Q93ul4 rhizobium l
29	7	5.0	283	16 Q8DID3	Q8did3 synecococc
30	7	5.0	283	17 Q8U2B4	Q8u2b4 pyrococcus
31	7	5.0	286	2 Q93UK9	Q93uk9 rhizobium e
32	7	5.0	290	2 Q93UL2	Q93ul2 rhizobium g
33	7	5.0	297	12 Q56780	Q56780 european ba
34	7	5.0	308	3 Q9CIK6	Q9cl6k neuropept
35	7	5.0	311	2 Q93UK6	Q93uk6 rhizobium l
36	7	5.0	323	16 Q8XCX9	Q8xcx9 salmonella
37	7	5.0	330	16 Q9I4V9	Q9i4v9 pseudomonas
38	7	5.0	331	16 Q9AD29	Q9ad29 streptomyce
39	7	5.0	332	6 Q8HXX3	Q8hxx3 macaca fasc
40	7	5.0	332	17 Q97YG7	Q97yg7 sulfolobus
41	7	5.0	338	2 Q8RPZ8	Q8rpz8 azotobacter
42	7	5.0	339	5 Q18567	Q18567 caenorhabdi
43	7	5.0	342	17 Q8TKF9	Q8tkf9 methanosaer
44	7	5.0	378	16 Q92DA8	Q92da8 listeria in
45	7	5.0	378	16 Q8Y8J3	Q8y8j3 listeria mo
46	7	5.0	391	6 Q95MNO	Q95mno cryptolagus
47	7	5.0	399	2 Q8RJM6	Q8rjm6 vibrio chol
48	7	5.0	409	16 P95182	P95182 mycobacteri
49	7	5.0	409	16 Q53957	Q53957 mycobacteri
50	7	5.0	410	2 Q99Q11	Q99q11 mycobacteri
51	7	5.0	422	17 Q8TVI6	Q8tvi6 methanopyru
52	7	5.0	423	16 Q8EJP7	Q8ejp7 shewanella
53	7	5.0	427	5 Q8MP98	Q8mp98 dictyosteli
54	7	5.0	443	2 Q8KLG3	Q8klg3 rhizobium e
55	7	5.0	456	16 Q8XPX8	Q8xpx8 ralstonia s
56	7	5.0	473	5 Q94416	Q94416 caenorhabdi
57	7	5.0	475	17 Q8UZY3	Q8uzy3 pyrococcus
58	7	5.0	490	16 Q8D3L6	Q8d3l6 vibrio vuln
59	7	5.0	491	16 Q9I2I8	Q9i2i8 pseudomonas
60	7	5.0	494	10 Q8RYN6	Q8ryn6 oryza sativ
61	7	5.0	496	5 Q8SWF5	Q8swf5 encephalito
62	7	5.0	499	5 Q966L0	Q966l0 caenorhabdi
63	7	5.0	512	2 Q87636	Q87636 pseudomonas
64	7	5.0	512	5 Q9V3A8	Q9v3a8 drosophila
65	7	5.0	519	16 Q9RKB5	Q9rkb5 streptomyce
66	7	5.0	524	16 Q53294	Q53294 mycobacteri
67	7	5.0	527	16 Q9I3H5	Q9i3h5 pseudomonas
68	7	5.0	538	3 Q43011	Q43011 schistosacch
69	7	5.0	591	2 Q8GNK9	Q8gnk9 mycobacteri
70	7	5.0	614	16 Q8FWB2	Q8fwb2 bruceella su
71	7	5.0	616	5 Q9BN24	Q9bn24 leishmania
72	7	5.0	637	16 Q8D3W6	Q8d3w6 vibrio vuln
73	7	5.0	670	8 Q9G927	Q9g927 ochronoxas
74	7	5.0	672	13 Q9W757	Q9w757 brachydanio
75	7	5.0	677	16 Q8PD55	Q8pd55 xanthomonas
76	7	5.0	687	13 Q9IAL0	Q9ial0 brachydanio
77	7	5.0	688	13 Q9W7R7	Q9w7r7 xenopus lae
78	7	5.0	688	13 Q9IBCB	Q9ibcb xenopus lae
79	7	5.0	694	16 Q8D5D8	Q8d5d8 rhizobium l
80	7	5.0	701	16 Q8PGQ5	Q8pgq5 xanthomonas
81	7	5.0	716	10 Q9SP26	Q9sp26 pisum sativ
82	7	5.0	731	16 Q8ZS53	Q8zss3 anabaena sp
83	7	5.0	747	3 Q8TFX3	Q8tfx3 aspergillus
84	7	5.0	748	16 Q8X6K3	Q8x6k3 escherichia
85	7	5.0	748	16 Q8FEB0	Q8feb0 escherichia
86	7	5.0	753	5 Q44861	Q44861 caenorhabdi
87	7	5.0	773	5 Q8IB77	Q8ib77 plasmodium
88	7	5.0	810	11 Q9ESZ9	Q9esz9 mus musculu
89	7	5.0	841	10 Q64542	Q64542 arabidopsis

90	7	5.0	860	10	Q8LIH0	Q8LIH0 oryza sativ	163	6	4.3	112	12	Q91IB3	Q91IB3 influenza a
91	7	5.0	860	11	Q8C9L8	Q8C9L8 mus musculus	164	6	4.3	112	12	Q91IB0	Q91IB0 influenza a
92	7	5.0	869	10	Q9FTV4	Q9FTV4 oryza sativ	165	6	4.3	112	12	Q91IA8	Q91IA8 influenza a
93	7	5.0	926	8	Q9GBS2	Q9GBS2 flammulina	166	6	4.3	113	16	Q8CKG0	Q8CKG0 yersinia pe
94	7	5.0	926	11	Q8R554	Q8R554 mus musculus	167	6	4.3	116	2	Q9RQI9	Q9RQI9 listeria m
95	7	5.0	964	16	Q8RJKC	Q8RJKC xanthomonas	168	6	4.3	116	16	Q9KNR2	Q9KNR2 aeropyrum p
96	7	5.0	1088	10	Q9WBC2	Q9WBC2 hordeum vul	169	6	4.3	116	17	Q9Y9Z9	Q9Y9Z9 vibrio chol
97	7	5.0	1175	17	Q8T123	Q8T123 methanosarc	170	6	4.3	117	16	Q9ACS9	Q9ACS9 streptomyce
98	7	5.0	1175	17	Q8PY33	Q8PY33 methanosarc	171	6	4.3	117	16	Q8DI75	Q8DI75 synchococc
99	7	5.0	1184	16	Q83909	Q83909 treponema p	172	6	4.3	117	16	Q8DI75	Q8DI75 synchococc
100	7	5.0	1203	5	Q8T1J7	Q8T1J7 dictyosteli	173	6	4.3	118	5	Q9NJH4	Q9NJH4 typanosoma
101	7	5.0	1218	5	Q96OH7	Q96OH7 drosophila	174	6	4.3	119	5	Q9VPT0	Q9VPT0 drosophila
102	7	5.0	1387	5	Q8IQC8	Q8IQC8 drosophila	175	6	4.3	120	16	Q8FI08	Q8FI08 leptospira
103	7	5.0	1392	5	Q9V745	Q9V745 drosophila	176	6	4.3	121	17	Q9BTB7	Q9BTB7 homo sapien
104	7	5.0	1547	5	Q9U1790	Q9U1790 caenorhabdi	177	6	4.3	122	4	Q9MTB7	Q9MTB7 homo sapien
105	7	5.0	2075	5	Q8IHR5	Q8IHR5 plasmodium	178	6	4.3	122	8	Q9MTI9	Q9MTI9 oenothera h
106	7	5.0	2422	16	Q8PI00	Q8PI00 xanthomonas	179	6	4.3	122	16	Q8Z599	Q8Z599 yersinia pe
107	7	5.0	2833	5	Q9VP13	Q9VP13 drosophila	180	6	4.3	123	12	Q8O842	Q8O842 hantavirus
108	7	5.0	2934	5	Q95XG5	Q95XG5 dictyosteli	181	6	4.3	123	16	Q9PGV5	Q9PGV5 xylella fas
109	7	5.0	4385	5	Q8I062	Q8I062 caenorhabdi	182	6	4.3	123	17	Q96XX0	Q96XX0 sulfolobus
110	7	5.0	4498	5	Q9WZ23	Q9WZ23 drosophila	183	6	4.3	125	10	Q942X3	Q942X3 oryza sativ
111	6	4.3	29	9	Q9XJN5	Q9XJN5 bacterioph	184	6	4.3	126	9	Q38151	Q38151 bacterioph
112	6	4.3	34	2	Q9RSB3	Q9RSB3 streptomyce	185	6	4.3	128	8	Q35408	Q35408 perga conde
113	6	4.3	38	9	Q9XJN1	Q9XJN1 bacterioph	186	6	4.3	128	9	Q34067	Q34067 streptococ
114	6	4.3	45	12	Q68607	Q68607 hepatitis c	187	6	4.3	128	9	Q8O175	Q8O175 streptococ
115	6	4.3	54	5	Q9U7K4	Q9U7K4 plasmodium	188	6	4.3	128	11	Q9QDQ9	Q9QDQ9 mus musculu
116	6	4.3	57	17	Q8TZH7	Q8TZH7 pyrococcus	189	6	4.3	128	11	Q9QDQ9	Q9QDQ9 mus musculu
117	6	4.3	59	5	Q9BUT8	Q9BUT8 plasmodium	190	6	4.3	128	16	Q06412	Q06412 mycobacteri
118	6	4.3	59	5	Q9WPH3	Q9WPH3 equine herp	191	6	4.3	129	4	Q9HAI2	Q9HAI2 homo sapien
119	6	4.3	59	12	Q8VJ64	Q8VJ64 mycobacteri	192	6	4.3	130	10	Q8GWM5	Q8GWM5 arabidopsis
120	6	4.3	60	12	Q8VJ64	Q8VJ64 canine herp	193	6	4.3	130	12	Q8JQF5	Q8JQF5 equine herp
121	6	4.3	60	12	Q96699	Q96699 canine herp	194	6	4.3	130	15	Q9S2S7	Q9S2S7 rhizobium m
122	6	4.3	60	12	Q96744	Q96744 feline herp	195	6	4.3	131	17	Q9HSY0	Q9HSY0 halobacteri
123	6	4.3	60	12	Q8BAH2	Q8BAH2 lung-eye-tr	196	6	4.3	131	17	Q8Z2P3	Q8Z2P3 pyrobaculum
124	6	4.3	62	16	Q8X4E7	Q8X4E7 escherichia	197	6	4.3	132	4	Q8H8K5	Q8H8K5 homo sapien
125	6	4.3	63	6	Q8MJ53	Q8MJ53 equus cabal	198	6	4.3	133	5	Q8MZ97	Q8MZ97 raietonia s
126	6	4.3	63	12	Q9QCV1	Q9QCV1 bottlenose	199	6	4.3	135	5	Q8V936	Q8V936 drosophila
127	6	4.3	67	2	Q93PX7	Q93PX7 pseudomonas	200	6	4.3	135	5	Q9V936	Q9V936 drosophila
128	6	4.3	67	16	Q9KOR1	Q9KOR1 neisseria m	201	6	4.3	135	16	Q9ABQ0	Q9ABQ0 caulobacter
129	6	4.3	71	2	Q93M44	Q93M44 bordetella	202	6	4.3	137	16	Q9ABP9	Q9ABP9 caulobacter
130	6	4.3	73	2	Q8KN06	Q8KN06 bacillus th	203	6	4.3	141	4	Q9Y425	Q9Y425 homo sapien
131	6	4.3	75	16	Q928Y4	Q928Y4 listeria in	204	6	4.3	141	9	Q94M65	Q94M65 streptococ
132	6	4.3	76	4	Q96EN4	Q96EN4 homo sapien	205	6	4.3	142	11	Q920V9	Q920V9 mus musculu
133	6	4.3	77	5	Q91945	Q91945 caenorhabdi	206	6	4.3	143	11	Q9D9P5	Q9D9P5 mus musculu
134	6	4.3	78	16	Q93M94	Q93M94 clostridium	207	6	4.3	143	16	Q9A540	Q9A540 caulobacter
135	6	4.3	80	16	Q8VJ99	Q8VJ99 mycobacteri	208	6	4.3	143	16	Q9AWC1	Q9AWC1 agrobacteri
136	6	4.3	82	10	Q8O943	Q8O943 arabidopsis	209	6	4.3	145	5	Q61351	Q61351 drosophila
137	6	4.3	86	16	Q9AGE8	Q9AGE8 caulobacter	210	6	4.3	145	5	Q9Y122	Q9Y122 hydra magni
138	6	4.3	89	15	Q9KUX5	Q9KUX5 vibrio chol	211	6	4.3	145	9	Q38182	Q38182 lactococcus
139	6	4.3	90	15	Q66735	Q66735 equine infe	212	6	4.3	145	9	Q9B014	Q9B014 bacterioph
140	6	4.3	91	5	Q9ENM6	Q9ENM6 hiltonius s	213	6	4.3	145	16	Q9CGT2	Q9CGT2 lactococcus
141	6	4.3	91	5	Q9ENM6	Q9ENM6 hiltonius s	214	6	4.3	146	3	Q9P492	Q9P492 podospira a
142	6	4.3	95	10	Q9LDM6	Q9LDM6 oryza sativ	215	6	4.3	146	3	Q9P492	Q9P492 podospira a
143	6	4.3	96	15	Q8USX3	Q8USX3 human immun	216	6	4.3	146	3	Q9P492	Q9P492 podospira a
144	6	4.3	97	16	Q8NQS2	Q8NQS2 corynebacte	217	6	4.3	147	5	Q8T176	Q8T176 bacterioph
145	6	4.3	98	16	Q8RPF1	Q8RPF1 fusobacteri	218	6	4.3	148	5	Q9GTU0	Q9GTU0 dictyosteli
146	6	4.3	100	12	Q91LLA	Q91LLA white spot	219	6	4.3	148	10	Q9LXH1	Q9LXH1 arabidopsis
147	6	4.3	101	10	Q8VYX7	Q8VYX7 pisum sativ	220	6	4.3	148	16	Q8Z0Q3	Q8Z0Q3 anabaena sp
148	6	4.3	101	17	Q9Y8V3	Q9Y8V3 aeropyrum p	221	6	4.3	149	3	Q8TFM2	Q8TFM2 aspergillus
149	6	4.3	102	16	Q8UI56	Q8UI56 agrobacteri	222	6	4.3	149	6	Q95JR2	Q95JR2 macaca fasc
150	6	4.3	103	4	Q9PIG0	Q9PIG0 homo sapien	223	6	4.3	149	9	Q9XJN9	Q9XJN9 bacterioph
151	6	4.3	103	17	Q9V032	Q9V032 pyrococcus	224	6	4.3	151	2	Q9RQ44	Q9RQ44 mycoplasma
152	6	4.3	104	2	Q9EY10	Q9EY10 streptomyce	225	6	4.3	151	5	Q9SVQ0	Q9SVQ0 microthera y
153	6	4.3	104	16	Q9WYV7	Q9WYV7 thermotoga	226	6	4.3	151	8	Q9G622	Q9G622 microthera y
154	6	4.3	104	16	Q9CD46	Q9CD46 mycobacteri	227	6	4.3	151	10	Q9FFY3	Q9FFY3 arabidopsis
155	6	4.3	105	2	Q9AEZ3	Q9AEZ3 frankia sp.	228	6	4.3	151	13	Q90XG5	Q90XG5 prototenus
156	6	4.3	105	5	Q8SXZ5	Q8SXZ5 drosophila	229	6	4.3	152	4	Q14713	Q14713 homo sapien
157	6	4.3	107	15	Q8USY0	Q8USY0 human immun	230	6	4.3	152	11	Q8C2P0	Q8C2P0 mus musculu
158	6	4.3	108	16	Q8DLK4	Q8DLK4 synchococc	231	6	4.3	153	16	Q8E390	Q8E390 streptococ
159	6	4.3	111	3	Q04293	Q04293 pichia angu	232	6	4.3	153	16	Q8DXM0	Q8DXM0 oryza sativ
160	6	4.3	112	5	Q9NFG9	Q9NFG9 plasmodium	233	6	4.3	154	10	Q9XEU8	Q9XEU8 hepatitis c
161	6	4.3	112	12	Q91IB1	Q91IB1 influenza a	234	6	4.3	154	12	Q917C2	Q917C2 vibrio chol
162	6	4.3	112	12	Q91IB2	Q91IB2 influenza a	235	6	4.3	154	16	Q86495	Q86495 streptomyce

236	6	4.3	155	2	Q9WWT8	Q9WWT8 pseudomonas	309	6	4.3	188	11	Q04365	Q04365 mus musculus
237	6	4.3	156	2	Q9FBN3	Q9FBN3 carboxydoth	310	6	4.3	188	13	Q90350	Q90350 coturnix co
238	6	4.3	156	4	Q96LR6	Q96LR6 homo sapien	311	6	4.3	188	16	Q83722	Q83722 treponema p
239	6	4.3	156	12	Q66089	Q66089 canine herp	312	6	4.3	188	16	Q8CR07	Q8CR07 staphylococ
240	6	4.3	157	2	Q93R95	Q93R95 thermus the	313	6	4.3	189	16	Q8NPP3	Q8NPP3 corynebacte
241	6	4.3	157	11	Q8BKAE	Q8BKAE mus musculus	314	6	4.3	190	2	Q8KTW0	Q8KTW0 escherichia
242	6	4.3	157	2	Q9F018	Q9F018 streptococc	315	6	4.3	190	16	Q8DIY7	Q8DIY7 synchococc
243	6	4.3	158	10	Q8QW99	Q8QW99 arabidopsis	316	6	4.3	191	2	Q9RFS2	Q9RFS2 vibrio vuln
244	6	4.3	158	16	Q91206	Q91206 pseudomonas	317	6	4.3	192	2	Q8KLY5	Q8KLY5 pseudomonas
245	6	4.3	158	16	Q97SD8	Q97SD8 streptococc	318	6	4.3	192	2	Q8KLY5	Q8KLY5 pseudomonas
246	6	4.3	159	16	Q92NG2	Q92NG2 staphylococ	319	6	4.3	192	2	Q8KLY5	Q8KLY5 pseudomonas
247	6	4.3	159	16	Q8CPV2	Q8CPV2 staphylococ	320	6	4.3	192	2	Q8KLY5	Q8KLY5 pseudomonas
248	6	4.3	160	5	Q81TA7	Q81TA7 aequiptecten	321	6	4.3	192	5	Q8WPV7	Q8WPV7 plasmodium
249	6	4.3	160	16	Q9X002	Q9X002 thermotoga	322	6	4.3	192	16	Q8XUS3	Q8XUS3 ralstonia s
250	6	4.3	160	16	Q8FMV3	Q8FMV3 corynebacte	323	6	4.3	193	2	Q8KUH9	Q8KUH9 actinosynne
251	6	4.3	161	11	Q9QZG0	Q9QZG0 rattus norv	324	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
252	6	4.3	162	11	Q8BTI3	Q8BTI3 mus musculus	325	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
253	6	4.3	162	16	Q8F4U1	Q8F4U1 leptospira	326	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
254	6	4.3	162	17	Q8F673	Q8F673 pyrococcus	327	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
255	6	4.3	163	2	Q9REV3	Q9REV3 variovorax	328	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
256	6	4.3	163	5	Q8WQ90	Q8WQ90 caenorhabdi	329	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
257	6	4.3	164	16	Q9Z582	Q9Z582 streptomyc	330	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
258	6	4.3	164	17	Q9ZUF3	Q9ZUF3 pyrococcus	331	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
259	6	4.3	165	4	Q96PC6	Q96PC6 homo sapien	332	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
260	6	4.3	165	16	Q9RS29	Q9RS29 deinococcus	333	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
261	6	4.3	166	5	Q18971	Q18971 caenorhabdi	334	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
262	6	4.3	166	10	Q8RZ96	Q8RZ96 oryza sativ	335	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
263	6	4.3	166	16	Q8DR04	Q8DR04 streptococc	336	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
264	6	4.3	167	1	Q9VZ23	Q9VZ23 methanococ	337	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
265	6	4.3	167	2	Q9X7I3	Q9X7I3 prochloroco	338	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
266	6	4.3	169	2	Q8KRW5	Q8KRW5 rhizobium l	339	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
267	6	4.3	169	11	Q9D5G1	Q9D5G1 mus musculus	340	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
268	6	4.3	170	12	Q9J3M1	Q9J3M1 human rotav	341	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
269	6	4.3	171	10	Q8LKF4	Q8LKF4 malus domes	342	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
270	6	4.3	171	12	Q91IR8	Q91IR8 hepatitis c	343	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
271	6	4.3	171	12	Q91IR7	Q91IR7 hepatitis c	344	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
272	6	4.3	171	12	Q91IS7	Q91IS7 hepatitis c	345	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
273	6	4.3	171	12	Q91IS2	Q91IS2 hepatitis c	346	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
274	6	4.3	171	12	Q91IR9	Q91IR9 hepatitis c	347	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
275	6	4.3	171	12	Q91IS6	Q91IS6 hepatitis c	348	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
276	6	4.3	172	16	Q9A3B5	Q9A3B5 caulobacter	349	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
277	6	4.3	172	16	Q9A3B5	Q9A3B5 rhizobium m	350	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
278	6	4.3	173	5	Q8MNI7	Q8MNI7 dictyosteli	351	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
279	6	4.3	174	10	Q22084	Q22084 petunia hyb	352	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
280	6	4.3	176	16	Q9AB29	Q9AB29 caulobacter	353	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
281	6	4.3	177	16	Q8FA24	Q8FA24 xanthomonas	354	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
282	6	4.3	178	2	Q33538	Q33538 rhodococcus	355	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
283	6	4.3	178	10	Q9SVD6	Q9SVD6 arabidopsis	356	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
284	6	4.3	178	16	P74585	P74585 synchocyst	357	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
285	6	4.3	179	16	Q8BZX2	Q8BZX2 leptospira	358	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
286	6	4.3	181	4	Q9H574	Q9H574 homo sapien	359	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
287	6	4.3	181	4	Q9BUC7	Q9BUC7 homo sapien	360	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
288	6	4.3	181	11	Q8K249	Q8K249 mus musculus	361	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
289	6	4.3	181	11	Q8BGR0	Q8BGR0 mus musculus	362	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
290	6	4.3	181	16	Q8F915	Q8F915 leptospira	363	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
291	6	4.3	182	5	Q9XZV0	Q9XZV0 leishmania	364	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
292	6	4.3	182	8	Q95H21	Q95H21 octoblephar	365	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
293	6	4.3	182	13	Q8QGC6	Q8QGC6 xenopus lae	366	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
294	6	4.3	182	16	Q8EKP8	Q8EKP8 shewanella	367	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
295	6	4.3	182	17	Q26796	Q26796 methanobact	368	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
296	6	4.3	183	2	Q69302	Q69302 campylobact	369	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
297	6	4.3	183	16	Q9PNM3	Q9PNM3 campylobact	370	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
298	6	4.3	183	16	Q911Y7	Q911Y7 pseudomonas	371	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
299	6	4.3	184	16	Q8ERM9	Q8ERM9 oceanobacil	372	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
300	6	4.3	184	17	Q8TLV9	Q8TLV9 methanosarc	373	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
301	6	4.3	186	10	Q9LFD2	Q9LFD2 arabidopsis	374	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
302	6	4.3	186	17	Q8PSA3	Q8PSA3 methanosarc	375	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
303	6	4.3	187	4	Q96DT9	Q96DT9 homo sapien	376	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
304	6	4.3	187	5	Q8IRM1	Q8IRM1 drosophila	377	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
305	6	4.3	187	11	Q9MUZ5	Q9MUZ5 mus musculus	378	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
306	6	4.3	187	16	Q8X3S7	Q8X3S7 escherichia	379	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
307	6	4.3	188	4	Q96T59	Q96T59 homo sapien	380	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
308	6	4.3	188	4	Q96T59	Q96T59 homo sapien	381	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
309	6	4.3	188	13	Q90350	Q90350 coturnix co	310	6	4.3	188	13	Q90350	Q90350 coturnix co
310	6	4.3	188	16	Q83722	Q83722 treponema p	311	6	4.3	188	16	Q83722	Q83722 treponema p
311	6	4.3	188	16	Q8CR07	Q8CR07 staphylococ	312	6	4.3	188	16	Q8CR07	Q8CR07 staphylococ
312	6	4.3	189	16	Q8NPP3	Q8NPP3 corynebacte	313	6	4.3	189	16	Q8NPP3	Q8NPP3 corynebacte
313	6	4.3	190	2	Q8KTW0	Q8KTW0 escherichia	314	6	4.3	190	2	Q8KTW0	Q8KTW0 escherichia
314	6	4.3	190	16	Q8DIY7	Q8DIY7 synchococc	315	6	4.3	190	16	Q8DIY7	Q8DIY7 synchococc
315	6	4.3	191	2	Q9RFS2	Q9RFS2 vibrio vuln	316	6	4.3	191	2	Q9RFS2	Q9RFS2 vibrio vuln
316	6	4.3	192	2	Q8KLY5	Q8KLY5 pseudomonas	317	6	4.3	192	2	Q8KLY5	Q8KLY5 pseudomonas
317	6	4.3	192	2	Q8KLY5	Q8KLY5 pseudomonas	318	6	4.3	192	2	Q8KLY5	Q8KLY5 pseudomonas
318	6	4.3	192	5	Q8WPV7	Q8WPV7 plasmodium	319	6	4.3	192	5	Q8WPV7	Q8WPV7 plasmodium
319	6	4.3	193	2	Q8XUS3	Q8XUS3 ralstonia s	320	6	4.3	193	2	Q8XUS3	Q8XUS3 ralstonia s
320	6	4.3	193	16	Q8KUH9	Q8KUH9 actinosynne	321	6	4.3	193	16	Q8KUH9	Q8KUH9 actinosynne
321	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s	322	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
322	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s	323	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
323	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s	324	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
324	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s	325	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
325	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s	326	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
326	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s	327	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
327	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s	328	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
328	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s	329	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
329	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s	330	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
330	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s	331	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
331	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s	332	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
332	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s	333	6	4.3	193	16	Q8XTI3	Q8XTI3 ralstonia s
333	6	4.3	193	16									

382	6	4.3	218	10	O24123	O24123 nicotiana s	455	6	4.3	245	16	Q9L170	Q9L170 streptomyce
383	6	4.3	218	11	Q9CX6	Q9CX6 mus musculus	456	6	4.3	246	10	Q8LN31	Q8LN31 oryza sativ
384	6	4.3	218	16	Q9ZB2	Q9ZB2 rickettsia	457	6	4.3	246	10	Q987P6	Q987P6 rhizobium l
385	6	4.3	218	16	Q9FGS	Q9FGS escherichia	458	6	4.3	246	16	Q8XGK1	Q8XGK1 salmonella
386	6	4.3	218	16	Q9D0C3	Q9D0C3 yersinia pe	459	6	4.3	246	16	Q8P4F1	Q8P4F1 xanthomonas
387	6	4.3	219	4	Q96P04	Q96P04 homo sapien	460	6	4.3	247	17	Q8TVU6	Q8TVU6 methanopyru
388	6	4.3	219	16	Q9HZA9	Q9HZA9 pseudomonas	461	6	4.3	248	4	Q96J10	Q96J10 homo sapien
389	6	4.3	221	4	Q9HCC6	Q9HCC6 homo sapien	462	6	4.3	248	12	Q919H9	Q919H9 culex nigri
390	6	4.3	221	5	Q95WZ5	Q95WZ5 ixodes scap	463	6	4.3	249	2	Q93H55	Q93H55 streptomyce
391	6	4.3	221	10	Q8H3V7	Q8H3V7 oryza sativ	464	6	4.3	249	2	Q50444	Q50444 mycobacteri
392	6	4.3	221	16	Q8XFR1	Q8XFR1 salmonella	465	6	4.3	250	16	Q9R9G9	Q9R9G9 homo sapien
393	6	4.3	221	16	Q8CJ3	Q8CJ3 streptomyce	466	6	4.3	250	16	Q92RJ5	Q92RJ5 rhizobium m
394	6	4.3	222	8	Q9WG12	Q9WG12 kongobathia	467	6	4.3	250	17	Q8ZV13	Q8ZV13 pyrobaculum
395	6	4.3	223	2	Q68496	Q68496 legionella	468	6	4.3	251	2	Q9FCX3	Q9FCX3 clostridium
396	6	4.3	223	10	Q8ZR3	Q8ZR3 oryza sativ	469	6	4.3	251	9	Q64113	Q64113 bacterioph
397	6	4.3	223	10	Q93X98	Q93X98 antirrhinum	470	6	4.3	251	16	Q34449	Q34449 bacillus su
398	6	4.3	224	8	Q8M834	Q8M834 microstomum	471	6	4.3	252	10	Q9S2Z6	Q9S2Z6 arabisopsis
399	6	4.3	225	4	Q96PJ1	Q96PJ1 homo sapien	472	6	4.3	252	10	Q8LIX6	Q8LIX6 oryza sativ
400	6	4.3	225	6	Q9BFM2	Q9BFM2 tapirus ind	473	6	4.3	252	17	Q9Y984	Q9Y984 aeropyrum p
401	6	4.3	225	6	Q9BFM4	Q9BFM4 equus cabal	474	6	4.3	253	10	Q8LHM5	Q8LHM5 oryza sativ
402	6	4.3	225	6	Q9BFM3	Q9BFM3 ceratotheri	475	6	4.3	253	10	Q40899	Q40899 petunia hyb
403	6	4.3	225	16	Q9PIL2	Q9PIL2 xanthomonas	476	6	4.3	253	16	Q929R2	Q929R2 listeria in
404	6	4.3	225	16	Q8P793	Q8P793 xanthomonas	477	6	4.3	253	16	Q8Y5F7	Q8Y5F7 listeria mo
405	6	4.3	225	17	Q8PXX5	Q8PXX5 sulfobolus	478	6	4.3	253	16	Q8P6C1	Q8P6C1 xanthomonas
406	6	4.3	225	17	Q8ZWB2	Q8ZWB2 pyrobaculum	479	6	4.3	253	16	Q9FC93	Q9FC93 streptomyce
407	6	4.3	226	12	Q66123	Q66123 cowpea mott	480	6	4.3	253	17	Q8TRU6	Q8TRU6 methanosarc
408	6	4.3	226	16	Q8Z823	Q8Z823 chlamydia p	481	6	4.3	253	17	Q8PV49	Q8PV49 methanosarc
409	6	4.3	226	16	Q8Z8J1	Q8Z8J1 vibrio vuln	482	6	4.3	254	5	Q5M012	Q5M012 drosophila
410	6	4.3	226	17	Q8Z841	Q8Z841 methanobact	483	6	4.3	255	2	Q8VR78	Q8VR78 escherichia
411	6	4.3	227	16	Q9CF48	Q9CF48 lactococcus	484	6	4.3	255	10	Q9S2Z4	Q9S2Z4 arabisopsis
412	6	4.3	228	2	Q9EUT0	Q9EUT0 staphylococ	485	6	4.3	255	16	Q9L114	Q9L114 streptomyce
413	6	4.3	228	10	Q94CV1	Q94CV1 oryza sativ	486	6	4.3	255	16	Q8CMG6	Q8CMG6 escherichia
414	6	4.3	228	16	Q99U36	Q99U36 staphylococ	487	6	4.3	255	17	Q8TZ29	Q8TZ29 pyrococcus
415	6	4.3	229	10	Q94P05	Q94P05 arabisopsis	488	6	4.3	255	13	Q9PUY7	Q9PUY7 osteglossu
416	6	4.3	229	10	Q8H0X7	Q8H0X7 arabisopsis	489	6	4.3	256	16	Q8CPH6	Q8CPH6 staphylococ
417	6	4.3	229	16	Q9PDK5	Q9PDK5 xylella fas	490	6	4.3	257	5	Q97250	Q97250 plasmodium
418	6	4.3	230	10	Q9FP31	Q9FP31 oryza sativ	491	6	4.3	257	16	Q92N06	Q92N06 rhizobium m
419	6	4.3	230	16	Q8EEH9	Q8EEH9 shewanella	492	6	4.3	258	4	Q9BU20	Q9BU20 homo sapien
420	6	4.3	231	10	Q8H555	Q8H555 oryza sativ	493	6	4.3	258	11	Q9QXV0	Q9QXV0 mus musculu
421	6	4.3	231	16	Q8NQD2	Q8NQD2 corynebacte	494	6	4.3	258	11	Q91W26	Q91W26 mus musculu
422	6	4.3	232	2	Q8KPR1	Q8KPR1 synchococ	495	6	4.3	259	2	Q303094	Q303094 streptomyce
423	6	4.3	232	2	Q8KQ9	Q8KQ9 rhizobium l	496	6	4.3	259	2	Q9K7D6	Q9K7D6 bacillus ha
424	6	4.3	232	10	Q9FVW9	Q9FVW9 arabisopsis	497	6	4.3	259	16	Q9UHG2	Q9UHG2 homo sapien
425	6	4.3	232	16	Q8FCX6	Q8FCX6 brucella me	498	6	4.3	260	4	Q9UHG2	Q9UHG2 rattus norv
426	6	4.3	232	16	Q8XRJ0	Q8XRJ0 raietonia s	499	6	4.3	260	11	Q9QXU9	Q9QXU9 corynebacte
427	6	4.3	233	17	Q97U09	Q97U09 sulfobolus	500	6	4.3	260	16	Q8NN19	Q8NN19 arabisopsis
428	6	4.3	234	2	Q85452	Q85452 rhizobium l	501	6	4.3	261	10	Q9LFD5	Q9LFD5 arabisopsis
429	6	4.3	234	16	Q8FVF1	Q8FVF1 brucella su	502	6	4.3	263	4	Q9BRU3	Q9BRU3 homo sapien
430	6	4.3	235	2	Q45235	Q45235 bradyrhizob	503	6	4.3	263	16	Q8XTA5	Q8XTA5 raietonia s
431	6	4.3	235	11	Q63349	Q63349 rattus norv	504	6	4.3	264	4	Q96HV5	Q96HV5 homo sapien
432	6	4.3	235	16	Q67273	Q67273 aquifex ae	505	6	4.3	264	4	Q8WX40	Q8WX40 homo sapien
433	6	4.3	235	16	Q9Z7T6	Q9Z7T6 chlamydia p	506	6	4.3	264	5	Q8T2L7	Q8T2L7 dictyosteli
434	6	4.3	235	16	Q8FHC0	Q8FHC0 escherichia	507	6	4.3	264	11	Q9D8U2	Q9D8U2 mus musculu
435	6	4.3	236	10	Q8S5W2	Q8S5W2 oryza sativ	508	6	4.3	265	2	Q46240	Q46240 clostridium
436	6	4.3	236	16	Q8YMB7	Q8YMB7 anabaena sp	509	6	4.3	265	4	Q43388	Q43388 homo sapien
437	6	4.3	236	17	Q27708	Q27708 methanobact	510	6	4.3	265	17	Q26265	Q26265 methanobact
438	6	4.3	237	5	Q9U2K4	Q9U2K4 caenorhabdi	511	6	4.3	266	16	Q98P77	Q98P77 rhizobium l
439	6	4.3	237	16	Q8UI25	Q8UI25 agrobacteri	512	6	4.3	267	2	Q54513	Q54513 yersinia en
440	6	4.3	237	16	Q9RUJ3	Q9RUJ3 deinococcus	513	6	4.3	267	12	Q99H25	Q99H25 helicoverp
441	6	4.3	238	16	Q8EMR9	Q8EMR9 oceanobacil	514	6	4.3	267	16	Q56949	Q56949 yersinia pe
442	6	4.3	239	16	Q9RS43	Q9RS43 deinococcus	515	6	4.3	267	16	Q8U9N5	Q8U9N5 agrobacteri
443	6	4.3	239	16	Q9F172	Q9F172 vibrio vuln	516	6	4.3	269	2	Q9XG10	Q9XG10 salmonella
444	6	4.3	239	16	Q8CV66	Q8CV66 oceanobacil	517	6	4.3	269	5	Q8ITC0	Q8ITC0 aequitecten
445	6	4.3	240	16	Q8X9X0	Q8X9X0 escherichia	518	6	4.3	269	16	Q92W32	Q92W32 rhizobium m
446	6	4.3	240	17	Q97YF7	Q97YF7 sulfobolus	519	6	4.3	270	4	Q9BVX0	Q9BVX0 homo sapien
447	6	4.3	241	2	Q8RLB6	Q8RLB6 comanonas a	520	6	4.3	270	4	Q96PJ2	Q96PJ2 homo sapien
448	6	4.3	241	3	Q01110	Q01110 kluyveromyc	521	6	4.3	270	10	Q9FTN3	Q9FTN3 oryza sativ
449	6	4.3	241	16	Q92R45	Q92R45 rhizobium m	522	6	4.3	270	16	Q98D50	Q98D50 rhizobium l
450	6	4.3	243	16	Q98E46	Q98E46 rhizobium l	523	6	4.3	271	16	Q8RE14	Q8RE14 fusbobacteri
451	6	4.3	243	16	Q50153	Q50153 mycobacteri	524	6	4.3	271	2	Q929R2	Q929R2 bacillus ha
452	6	4.3	244	2	Q47880	Q47880 escherichia	525	6	4.3	271	2	Q93UK7	Q93UK7 rhizobium g
453	6	4.3	244	4	Q9BR78	Q9BR78 homo sapien	526	6	4.3	271	16	Q8P3B5	Q8P3B5 xanthomonas
454	6	4.3	244	16	Q926S7	Q926S7 chlamydia p	527	6	4.3	271	16	Q8E8D9	Q8E8D9 shewanella



528	6	4.3	272	17	Q8T233	Q8T233 methanopyru	601	6	4.3	294	16	Q92JV1	Q92JV1 rhizobium m
529	6	4.3	275	5	Q8RMV9	Q8RMV9 bacillus an	602	6	4.3	295	13	Q9DEI4	Q9DEI4 anser anser
530	6	4.3	275	5	Q9U3J0	Q9U3J0 caenorhabdi	603	6	4.3	296	2	Q8F8S8	Q8F8S8 streptomyc
531	6	4.3	275	16	Q92YR7	Q92YR7 listeria in	604	6	4.3	296	2	Q8GHA7	Q8GHA7 streptomyc
532	6	4.3	275	16	Q8Y478	Q8Y478 listeria mo	605	6	4.3	296	10	Q9FW94	Q9FW94 oryza sativ
533	6	4.3	275	16	Q8P2P8	Q8P2P8 streptococ	606	6	4.3	296	16	Q8FJ86	Q8FJ86 xanthomonas
534	6	4.3	275	16	Q9ALJ3	Q9ALJ3 streptococ	607	6	4.3	296	16	Q8P6I7	Q8P6I7 xanthomonas
535	6	4.3	276	2	Q8KKY4	Q8KKY4 rhizobium e	608	6	4.3	297	16	Q8Z894	Q8Z894 chlamydia p
536	6	4.3	276	16	Q51675	Q51675 borrelia bu	609	6	4.3	297	16	Q8PH15	Q8PH15 xanthomonas
537	6	4.3	277	2	Q8RST7	Q8RST7 erwinia chr	610	6	4.3	298	16	Q8Y0Y7	Q8Y0Y7 ralstonia s
538	6	4.3	277	4	Q8WIH1	Q8WIH1 homo sapien	611	6	4.3	300	3	Q9C2R0	Q9C2R0 neurospora
539	6	4.3	277	16	Q98HE1	Q98HE1 rhizobium l	612	6	4.3	300	10	Q9AVG1	Q9AVG1 oryza sativ
540	6	4.3	278	2	Q8ETR2	Q8ETR2 corynebacte	613	6	4.3	300	16	Q9A3W1	Q9A3W1 caulobacter
541	6	4.3	278	16	Q8FS37	Q8FS37 corynebacte	614	6	4.3	300	17	Q9HQH2	Q9HQH2 halobacteri
542	6	4.3	279	10	Q94CV4	Q94CV4 oryza sativ	615	6	4.3	301	2	Q938B6	Q938B6 mycobacteri
543	6	4.3	280	10	Q9MALL	Q9MALL arabidopsis	616	6	4.3	302	13	Q9DDC5	Q9DDC5 meleagris g
544	6	4.3	280	16	Q984C7	Q984C7 rhizobium l	617	6	4.3	302	13	Q9DDB7	Q9DDB7 struthio ca
545	6	4.3	281	1	Q8J308	Q8J308 pyrococcus	618	6	4.3	302	16	Q92UE2	Q92UE2 rhizobium m
546	6	4.3	281	10	Q9M340	Q9M340 arabidopsis	619	6	4.3	302	16	Q8XPR4	Q8XPR4 ralstonia s
547	6	4.3	281	16	Q8ZS03	Q8ZS03 salmonella	620	6	4.3	304	9	Q92XK8	Q92XK8 bacterioph
548	6	4.3	281	16	Q8Z9P6	Q8Z9P6 salmonella	621	6	4.3	304	16	Q8X544	Q8X544 escherichia
549	6	4.3	281	16	Q8EBG5	Q8EBG5 shewanella	622	6	4.3	304	16	Q8CS13	Q8CS13 staphylococ
550	6	4.3	282	5	Q8WP37	Q8WP37 dendronept	623	6	4.3	305	2	Q45818	Q45818 chloroflexu
551	6	4.3	282	10	Q9FL72	Q9FL72 arabidopsis	624	6	4.3	305	2	Q87967	Q87967 streptomyc
552	6	4.3	282	10	Q8LEA8	Q8LEA8 arabidopsis	625	6	4.3	305	16	Q98D14	Q98D14 rhizobium l
553	6	4.3	282	13	Q9DDB6	Q9DDB6 sita europ	626	6	4.3	305	16	Q986U5	Q986U5 rhizobium l
554	6	4.3	282	16	Q9AA63	Q9AA63 caulobacter	627	6	4.3	306	16	Q914D4	Q914D4 xanthomonas
555	6	4.3	283	3	Q93337	Q93337 saccharomyc	628	6	4.3	306	16	Q8PLD8	Q8PLD8 xanthomonas
556	6	4.3	283	5	Q9VIP7	Q9VIP7 drosophila	629	6	4.3	306	17	Q96ZP7	Q96ZP7 sulfolobus
557	6	4.3	283	5	Q8KSF4	Q8KSF4 drosophila	630	6	4.3	306	17	Q8ZXS4	Q8ZXS4 pyrobaculum
558	6	4.3	283	16	Q988E7	Q988E7 rhizobium l	631	6	4.3	307	2	Q9F186	Q9F186 alcaligenes
559	6	4.3	283	17	Q8ZWS1	Q8ZWS1 pyrobaculum	632	6	4.3	307	2	Q8N4L8	Q8N4L8 homo sapien
560	6	4.3	285	16	Q92ZG2	Q92ZG2 rhizobium m	633	6	4.3	307	4	Q9N4K9	Q9N4K9 homo sapien
561	6	4.3	285	17	Q96731	Q96731 methanobact	634	6	4.3	307	4	Q96FI5	Q96FI5 homo sapien
562	6	4.3	286	16	Q9CNY1	Q9CNY1 pasteurella	635	6	4.3	307	10	Q9ASS3	Q9ASS3 arabidopsis
563	6	4.3	286	16	Q8Z1E0	Q8Z1E0 salmonella	636	6	4.3	308	16	Q8PGA3	Q8PGA3 xanthomonas
564	6	4.3	286	16	Q8KCC8	Q8KCC8 chlorobium	637	6	4.3	308	17	Q9HPL3	Q9HPL3 halobacteri
565	6	4.3	287	16	Q9KTK0	Q9KTK0 vibrio chol	638	6	4.3	309	12	Q93UL0	Q93UL0 rhizobium g
566	6	4.3	287	16	Q8YPL6	Q8YPL6 anabaena sp	639	6	4.3	309	12	Q8V6M8	Q8V6M8 halovirus h
567	6	4.3	287	16	Q8UIM0	Q8UIM0 agrobacteri	640	6	4.3	309	16	Q92SN5	Q92SN5 rhizobium m
568	6	4.3	287	17	Q95897	Q95897 sulfolobus	641	6	4.3	309	16	Q8D7V0	Q8D7V0 streptococ
569	6	4.3	288	10	Q9M1H2	Q9M1H2 arabidopsis	642	6	4.3	310	2	Q93UL1	Q93UL1 rhizobium g
570	6	4.3	288	16	Q9CHJ0	Q9CHJ0 lactococcus	643	6	4.3	310	10	Q8S4C4	Q8S4C4 zea mays m
571	6	4.3	288	16	Q99335	Q99335 mycobacteri	644	6	4.3	311	2	Q93UL6	Q93UL6 mesorhizobi
572	6	4.3	288	16	Q8DI77	Q8DI77 synecococ	645	6	4.3	311	13	Q9DDF0	Q9DDF0 geochelone
573	6	4.3	289	10	Q95663	Q95663 clostrerium	646	6	4.3	312	10	Q8LRM2	Q8LRM2 petunia hyb
574	6	4.3	289	16	Q9A529	Q9A529 caulobacter	647	6	4.3	312	11	Q9QUR4	Q9QUR4 mus musculu
575	6	4.3	289	16	Q97DN3	Q97DN3 clostridium	648	6	4.3	312	11	Q9R225	Q9R225 mus musculu
576	6	4.3	289	16	P96206	P96206 mycobacteri	649	6	4.3	312	16	Q8P4N5	Q8P4N5 xanthomonas
577	6	4.3	290	2	Q66013	Q66013 photobacter	650	6	4.3	313	2	Q8RNM4	Q8RNM4 legionella
578	6	4.3	290	4	Q9BSR0	Q9BSR0 homo sapien	651	6	4.3	313	2	Q86466	Q86466 xanthomonas
579	6	4.3	290	16	Q92SR4	Q92SR4 rhizobium m	652	6	4.3	313	11	Q8VF90	Q8VF90 mus musculu
580	6	4.3	290	16	Q92RP8	Q92RP8 vibrio vuln	653	6	4.3	313	11	Q82UY2	Q82UY2 rhizobium m
581	6	4.3	291	2	Q50170	Q50170 helicobacte	654	6	4.3	313	16	Q92UY2	Q92UY2 agrobacteri
582	6	4.3	291	2	P94824	P94824 helicobacte	655	6	4.3	313	16	Q8UEA2	Q8UEA2 agrobacteri
583	6	4.3	291	10	Q940W0	Q940W0 raphanus sa	656	6	4.3	313	16	Q8PCA2	Q8PCA2 xanthomonas
584	6	4.3	291	16	Q9A2C8	Q9A2C8 caulobacter	657	6	4.3	313	17	Q8Z2Y2	Q8Z2Y2 pyrobaculum
585	6	4.3	292	2	Q9K314	Q9K314 vibrio para	658	6	4.3	314	16	Q92MU3	Q92MU3 rhizobium m
586	6	4.3	292	2	Q9K349	Q9K349 vibrio para	659	6	4.3	315	5	Q9N4G5	Q9N4G5 caenorhabdi
587	6	4.3	292	2	Q9K309	Q9K309 vibrio para	660	6	4.3	315	15	Q8T0Q2	Q8T0Q2 thiosophila
588	6	4.3	292	2	Q9K310	Q9K310 vibrio para	661	6	4.3	315	16	Q984T5	Q984T5 rhizobium l
589	6	4.3	292	4	Q8WU56	Q8WU56 homo sapien	662	6	4.3	316	2	Q8KPR7	Q8KPR7 synecococ
590	6	4.3	292	10	Q93671	Q93671 hordeum vul	663	6	4.3	316	16	Q9PHB4	Q9PHB4 xylella fas
591	6	4.3	292	16	Q9ZKX6	Q9ZKX6 helicobacte	664	6	4.3	316	16	Q8UHI2	Q8UHI2 agrobacteri
592	6	4.3	292	16	Q9AK27	Q9AK27 streptomyc	665	6	4.3	316	16	Q8RA41	Q8RA41 thermoaer
593	6	4.3	292	16	Q8FRP4	Q8FRP4 corynebacte	666	6	4.3	317	16	Q9WTD4	Q9WTD4 thermotoga
594	6	4.3	292	16	Q8EUG8	Q8EUG8 mycoplasma	667	6	4.3	317	16	Q8DW20	Q8DW20 streptococ
595	6	4.3	293	5	Q9MAU7	Q9MAU7 drosophila	668	6	4.3	318	10	Q9LOA5	Q9LOA5 arabidopsis
596	6	4.3	293	10	Q23798	Q23798 triticum ae	669	6	4.3	320	16	Q8XTB7	Q8XTB7 ralstonia s
597	6	4.3	293	16	Q9PEB9	Q9PEB9 xylella fas	670	6	4.3	321	2	Q9AF12	Q9AF12 vibrio para
598	6	4.3	294	5	Q8SQ14	Q8SQ14 encephalito	671	6	4.3	321	10	Q93Z37	Q93Z37 arabidopsis
599	6	4.3	294	5	Q8SRG2	Q8SRG2 encephalito	672	6	4.3	322	5	Q45537	Q45537 caenorhabdi
600	6	4.3	294	10	Q81987	Q81987 hordeum vul	673	6	4.3	322	16	Q8FXH3	Q8FXH3 escherichia



820	364	4.3	6	820	Q8f2a8 leptospira	893	6	4.3	392	13	Q9VHA5	Q9Vha5 myxine glut
821	365	4.3	6	821	Q68776 versinia pe	894	6	4.3	392	13	Q8AVB5	Q8avb5 cryzias lat
822	365	4.3	6	822	Q8NEW4	895	6	4.3	393	16	Q928F7	Q928f7 listeria in
823	365	4.3	6	823	Q9KfJ4 bacillus ha	896	6	4.3	393	16	O08395	O08395 mycobacteri
824	366	4.3	6	824	Q917Q7 drosophila	897	6	4.3	393	17	Q97YH4	Q97Yh4 sulfolobus
825	366	4.3	6	825	Q8BX06 mus musculu	898	6	4.3	394	2	Q9F241	Q9f241 xanthomonas
826	367	4.3	6	826	Q9ev98 arabidopsis	899	6	4.3	394	2	Q9XB66	Q9xb66 amycolatops
827	367	4.3	6	827	Q9sb42 arabidopsis	900	6	4.3	394	4	Q9Y3W7	Q9y3w7 homo sapien
828	368	4.3	6	828	Q9SB42	901	6	4.3	394	6	O8H247	O8h247 felis silve
829	369	4.3	6	829	Q9Jrnl actinobacil	902	6	4.3	394	6	P72780	P72780 synchocyst
830	369	4.3	6	830	Q9sv45 arabidopsis	903	6	4.3	395	2	Q9ZEG8	Q9zeg8 mycobacteri
831	369	4.3	6	831	Q9a3u6 caulobacter	904	6	4.3	395	2	Q9RAJ8	Q9raj8 mycobacteri
832	369	4.3	6	832	Q9ugl7 agrobacteri	905	6	4.3	395	2	Q9AKV5	Q9akv5 mycobacteri
833	370	5	6	833	O44562 caenorhabdi	906	6	4.3	395	11	Q9JHE0	Q9jhe0 rattus norv
834	370	5	6	834	Q94404 caenorhabdi	907	6	4.3	395	17	Q97XP1	Q97xp1 sulfolobus
835	370	16	6	835	Q98HP3 rhizobium 1	908	6	4.3	395	17	O8TKY3	O8tky3 methanosarc
836	370	16	6	836	O8Ynk9 anabaena ep	909	6	4.3	395	17	O8TKY3	O8tky3 methanosarc
837	370	17	6	837	O29911 archaeoglob	910	6	4.3	396	2	O8L160	O8l160 myxococcus
838	371	2	6	838	O27282 methanobact	911	6	4.3	396	2	Q9KTH4	Q9kth4 rhizobium e
839	371	2	6	839	P86443 rhizobium m	912	6	4.3	396	4	Q96K21	Q96k21 homo sapien
840	371	2	6	840	Q8ghq5 pseudomonas	913	6	4.3	396	5	O15564	O15564 trichomonas
841	371	5	6	841	Q9U2J4 caenorhabdi	914	6	4.3	396	5	Q9VVD9	Q9vvd9 drosophila
842	371	10	6	842	O81286 arabidopsis	915	6	4.3	396	5	O25553	O25553 naegleria f
843	371	11	6	843	Q9qwx9 mus musculu	916	6	4.3	396	10	Q94122	Q94122 oryza sativ
844	371	11	6	844	Q9qwy4 mus musculu	917	6	4.3	396	11	O61669	O61669 mus musculu
845	371	11	6	845	Q9qwx8 mus musculu	918	6	4.3	396	13	O8JHW6	O8jhw6 fugu rubrip
846	371	11	6	846	O54849 mus musculu	919	6	4.3	396	16	O8RI68	O8ri68 fusbacteri
847	371	11	6	847	Q9qwy1 mus musculu	920	6	4.3	396	16	O8FM37	O8fm37 corynebacte
848	371	11	6	848	O88912 mus musculu	921	6	4.3	397	2	O9AGF0	O9agf0 mycobacteri
849	371	11	6	849	Q9qwy5 mus musculu	922	6	4.3	397	5	O8IEB4	O8ieb4 plasmodium
850	371	16	6	850	Q9QUJ6	923	6	4.3	397	11	O8VDC4	O8vdc4 mus musculu
851	371	17	6	851	Q9PNS1	924	6	4.3	397	11	O9CWK7	O9cwk7 mus musculu
852	371	17	6	852	Q9HSF1	925	6	4.3	397	11	O8CFD5	O8cfd5 mus musculu
853	372	16	6	853	Q8PH06	926	6	4.3	397	16	Q9PH91	Q9ph91 xylella fas
854	372	16	6	854	Q8P5L2	927	6	4.3	397	17	Q9V095	Q9v095 pyrococcus
855	374	10	6	855	Q9CSL2	928	6	4.3	398	3	Q9HED1	Q9hed1 neurospora
856	374	16	6	856	Q9X8Y8	929	6	4.3	398	11	O8CB15	O8cb15 mus musculu
857	374	17	6	857	Q97V53 sulfolobus	930	6	4.3	399	5	O8SQR3	O8sqr3 encephalito
858	374	17	6	858	Q8PZY2	931	6	4.3	399	9	O8W735	O8w735 bacteriopho
859	376	4	6	859	O8WXH3 homo sapien	932	6	4.3	399	10	O8LG40	O8lg40 arabidopsis
860	376	16	6	860	Q9S2T3 streptomyce	933	6	4.3	399	16	P74550	P74550 synchocyst
861	376	17	6	861	Q97XA4 sulfolobus	934	6	4.3	399	17	Q97V47	Q97v47 sulfolobus
862	377	10	6	862	Q9fnc6 arabidopsis	935	6	4.3	400	15	Q90KJ2	Q90kj2 human immun
863	377	16	6	863	Q97EW8 clostridium	936	6	4.3	401	16	Q97F24	Q97f24 clostridium
864	377	16	6	864	O8XSG2 ralstonia s	937	6	4.3	401	17	O8PXS8	O8pxs8 methanosarc
865	378	10	6	865	O8SVZ0 arabidopsis	938	6	4.3	402	16	O8XST0	O8xst0 escherichia
866	378	16	6	866	O8EVB1	939	6	4.3	402	16	O33312	O33312 mycobacteri
867	379	15	6	867	O8J4S1	940	6	4.3	403	4	Q96CX6	Q96cx6 homo sapien
868	380	10	6	868	O81760 arabidopsis	941	6	4.3	403	5	Q9VCZ2	Q9vcz2 drosophila
869	381	4	6	869	O8N022	942	6	4.3	403	5	O8MZG6	O8mzg6 drosophila
870	381	5	6	870	O58542	943	6	4.3	403	10	O9FHS6	O9fhs6 arabidopsis
871	381	17	6	871	O88542 pyrococcus	944	6	4.3	403	17	Q97YI2	Q97yi2 sulfolobus
872	383	4	6	872	Q96DW5 homo sapien	945	6	4.3	404	5	O20212	O20212 caenorhabdi
873	383	4	6	873	Q969Y6 homo sapien	946	6	4.3	404	10	Q941V4	Q941v4 oryza sativ
874	384	11	6	874	Q9EPB7	947	6	4.3	404	12	Q98161	Q98161 little cher
875	384	11	6	875	O08905 mus musculu	948	6	4.3	404	12	O40962	O40962 little cher
876	384	11	6	876	Q9esp4 rattus norv	949	6	4.3	404	16	Q9EWI4	Q9ewi4 streptomyce
877	384	16	6	877	Q9A7D1 caulobacter	950	6	4.3	405	2	O8KT72	O8kt72 photorhabdu
878	385	5	6	878	O45442 caenorhabdi	951	6	4.3	406	2	Q9RCB5	Q9rcb5 versinia ps
879	385	5	6	879	O8IPB3 drosophila	952	6	4.3	406	5	O9N4T4	O9n4t4 caenorhabdi
880	387	12	6	880	O39887 hepatitis b	953	6	4.3	406	10	Q9LYA9	Q9lya9 arabidopsis
881	387	16	6	881	Q9KV18 vibrio chol	954	6	4.3	406	10	O8LA93	O8la93 arabidopsis
882	388	2	6	882	O8KHU2	955	6	4.3	406	10	Q9LV70	Q9lv70 arabidopsis
883	388	16	6	883	O9K8K1 bacillus ha	956	6	4.3	407	2	O93AL9	O93al9 clostridium
884	388	16	6	884	O8FLY3	957	6	4.3	407	10	O22028	O22028 cyanidium c
885	389	3	6	885	Q54287 streptomyce	958	6	4.3	407	10	O8GVH3	O8gvh3 arabidopsis
886	389	3	6	886	Q9HGL1 schizosacch	959	6	4.3	407	16	O8RJI9	O8rji9 streptomyce
887	390	2	6	887	Q9RH97 salmonella	960	6	4.3	408	16	Q9CDJ3	Q9cdj3 lactococcus
888	390	5	6	888	Q9GYJ3 caenorhabdi	961	6	4.3	410	17	Q97XQ8	Q97xq8 sulfolobus
889	390	10	6	889	Q9SR08	962	6	4.3	410	2	O48368	O48368 mycobacteri
890	390	16	6	890	O8XS34 ralstonia s	963	6	4.3	410	3	Q12665	Q12665 penicillium
891	391	10	6	891	P92929 antithamnio	964	6	4.3	411	2	O48882	O48882 mycobacteri
892	391	17	6	892	Q97Y44 sulfolobus	965	6	4.3	411	4	O8MW77	O8mw77 homo sapien
	392	11	6		Q99KP5 mus musculu		6	4.3	411	5	Q9VX69	Q9vx69 drosophila

Q9nf81 leishmania  
 Q8sxg0 drosophila  
 Q8u9u2 agrobacteri  
 Q97v25 sulfobus  
 Q8ztr2 pyrobaculum  
 Q9hfm0 metarhizium  
 Q8fn18 corynebacte  
 Q9rjg0 streptomyc  
 Q8n836 homo sapien  
 Q57456 synechocyst  
 Q9zito escherichia  
 Q8kmw4 escherichia  
 Q8wv18 homo sapien  
 O15828 leishmania  
 O62370 caenorhabdi  
 O70629 mus musculu  
 O88691 mus musculu  
 Q8fc97 escherichia  
 Q9fsf2 nicotiana t  
 P72207 pseudomonas  
 Q8eg8 rhizobium l  
 Q8emt7 oceanobacil  
 Q8fvt2 bruceella su  
 Q94245 schizosacch  
 Q9vni39 drosophila  
 Q93119 antheraea p  
 Q9rn56 streptomyc  
 Q9etg0 corynebacte  
 Q9f826 micromonosp  
 Q8idh7 caenorhabdi  
 Q68733 equine infe  
 O67991 pseudomonas  
 Q921s7 mus musculu  
 Q91422 gallus gall  
 Q8ah28 human immun

ALIGNMENTS

RESULT 1  
 Q8MNM4  
 ID Q8MNM4 PRELIMINARY; PRT; 141 AA.  
 AC Q8MNM4;  
 DT 01-OCT-2002 (TREMBLrel. 22, Created)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE Putative vir15 protein.  
 GN VIR.  
 OS Babesia canis.  
 OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Babesiidae; Babesia.  
 OX NCBI\_TaxID=5867;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Drakulovski P., Carcy B., Moubri K., Carret C., Depoix D.,  
 RA Schettlers T.P.M., Gorenflot A.;  
 RT "An extrachromosomal dRNA from Babesia canis implicated in parasite  
 RT virulence.";  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ494862; CAD41951.1; -;  
 SQ SEQUENCE 141 AA; 15752 MW; B84419C12BFD7CD1 CRC64;  
 Query Match 100.0%; Score 141; DB 5; Length 141;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-136;  
 Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MESTSTTNFVAENRPTFGFTFDVNRALLRVKSSERLALAGMCHRVLPOTGASA 60  
 Db 1 MESTSTTNFVAENRPTFGFTFDVNRALLRVKSSERLALAGMCHRVLPOTGASA 60  
 QY 61 IAATVTPKASMKLPPRPQSTKSPRLSKIRENMKTIQESARVNHRLPEGHPLLE 120  
 Db 61 IAATVTPKASMKLPPRPQSTKSPRLSKIRENMKTIQESARVNHRLPEGHPLLE 120

QY 121 KRAEYFRLRLSLKSGVNRLL 141  
 Db 121 KRAEYFRLRLSLKSGVNRLL 141  
 RESULT 2  
 Q8MQH4  
 ID Q8MQH4 PRELIMINARY; PRT; 285 AA.  
 AC Q8MQH4;  
 DT 01-OCT-2002 (TREMBLrel. 22, Created)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE Putative vir32 protein.  
 GN VIR.  
 OS Babesia canis.  
 OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Babesiidae; Babesia.  
 OX NCBI\_TaxID=5867;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Drakulovski P., Carcy B., Moubri K., Carret C., Depoix D.,  
 RA Schettlers T.P.M., Gorenflot A.;  
 RT "An extrachromosomal dRNA from Babesia canis implicated in parasite  
 RT virulence.";  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ494862; CAD41950.1; -;  
 SQ SEQUENCE 285 AA; 32169 MW; 45ACEFEC7927243D CRC64;  
 Query Match 84.4%; Score 119; DB 5; Length 285;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-113;  
 Matches 119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MESTSTTNFVAENRPTFGFTFDVNRALLRVKSSERLALAGMCHRVLPOTGASA 60  
 Db 1 MESTSTTNFVAENRPTFGFTFDVNRALLRVKSSERLALAGMCHRVLPOTGASA 60  
 QY 61 IAATVTPKASMKLPPRPQSTKSPRLSKIRENMKTIQESARVNHRLPEGHPLL 119  
 Db 61 IAATVTPKASMKLPPRPQSTKSPRLSKIRENMKTIQESARVNHRLPEGHPLL 119

RESULT 3

Q9LHSL  
 ID Q9LHSL PRELIMINARY; PRT; 334 AA.  
 AC Q9LHSL;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE Genomic DNA, chromosome 5, TAC clone:K3D20.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Columbia;  
 RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP002031; BAA98204.1; -;  
 DR InterPro; IPR002965; P-rich extensin.  
 DR PRINTS; PR01217; PRICHEXTEN.  
 SQ SEQUENCE 334 AA; 37849 MW; 5DF169A85F1B5FA9 CRC64;  
 Query Match 5.7%; Score 8; DB 10; Length 334;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 59 SAIAATVT 66  
 Db 20 SAIAATVT 27

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 59 SAIAATVT 66  
 Db 23 SAIAATVT 30

RESULT 6  
 Q8EGU2 PRELIMINARY; PRT; 534 AA.

ID Q8EGU2  
 AC Q8EGU2  
 DT 01-MAR-2003 (TRENBLrel. 23, Created)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Glycogen synthase.  
 DE GLGA OR S01499.  
 OS Shewanella oneidensis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
 OC Alteromonadaceae; Shewanella.  
 OX NCBI\_TaxID=70863;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MR-1;  
 RX MEDLINE=22397686; PubMed=12368813;  
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,  
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,  
 RA Meyer T., Teapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,  
 RA Vamathevan J., Weidman J., Imbraim M., Lee K., Berry K., Lee C.,  
 RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,  
 RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.;  
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
 RT Shewanella oneidensis";  
 RL Nat. Biotechnol. 20:1118-1123(2002).  
 DR EMBL; AE015595; AAN54560.1; -.  
 DR TIGR; S01499; -.  
 KW Complete proteome.  
 SQ SEQUENCE 534 AA; 58933 MW; BB2662614BB6993 CRC64;

Query Match 5.7%; Score 8; DB 16; Length 534;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 LAMLRALA 45  
 Db 151 LAMLRALA 158

Query Match 5.7%; Score 8; DB 16; Length 534;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 7  
 Q9SK35 PRELIMINARY; PRT; 559 AA.

ID Q9SK35  
 AC Q9SK35  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
 DE At2G24980 protein.  
 GN AT2G24980.  
 OS Arabidopsis thaliana (Mouse-ear cross).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Renning C.W., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,  
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,

RA Salzberg S.L., Fraser C.M., Venter J.C.;  
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
thaliana";  
RL Nature 402:761-768 (1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Lin X.;  
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AC006585; AAD23015.1; -  
DR InterPro; IPR002965; P rich. extensn.  
DR PRINTS; PR01217; PRICHEXTENSN.  
SQ SEQUENCE 559 AA; 63164 MW; B22AA763D0D1CC6D CRC64;

Query Match 5.7%; Score 8; DB 10; Length 559;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 SAIAATVT 66  
| | | | |  
DB 14 SAIAATVT 21

RESULT 8  
Q9FH26 PRELIMINARY; PRT; 609 AA.  
ID Q9FH26;  
AC Q9FH26;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Genomic DNA, chromosome 5, TAC clone:K20J1.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Columbia;  
RX MEDLINE=20181125; PubMed=10718197;  
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,  
RA Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence  
features of the regions of 3,076,755 bp covered by sixty P1 and TAC  
clones";  
RL DNA Res. 7:31-63 (2000).  
DR EMBL; AB023028; BAB10090.1; -  
DR InterPro; IPR002965; P rich. extensn.  
DR PRINTS; PR01217; PRICHEXTENSN.  
SQ SEQUENCE 609 AA; 68454 MW; C963750F7539DBA5 CRC64;

Query Match 5.7%; Score 8; DB 10; Length 609;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 SAIAATVT 66  
| | | | |  
DB 14 SAIAATVT 21

RESULT 9  
Q9FG06 PRELIMINARY; PRT; 689 AA.  
ID Q9FG06;  
AC Q9FG06;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Genomic DNA, chromosome 5, BAC clone:F15W7.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Columbia;  
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,  
RA Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF002543; BAB11413.1; -  
DR InterPro; IPR002965; P rich. extensn.  
DR PRINTS; PR01217; PRICHEXTENSN.  
SQ SEQUENCE 689 AA; 77507 MW; 6AE991A17A2536C7 CRC64;

Query Match 5.7%; Score 8; DB 10; Length 689;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 SAIAATVT 66  
| | | | |  
DB 39 SAIAATVT 46

RESULT 10  
Q9STM7 PRELIMINARY; PRT; 707 AA.  
ID Q9STM7;  
AC Q9STM7;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Extensin-like protein.  
GN T28D5.100 OR AT4G08410.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bevan M., Lennard N., Quail M., Harris B., Rajandream M.A.,  
RA Barrell B.G., Bancroft I., Mewes H.W., Mayer K.F.X., Lemcke K.,  
RA Schueller C.;  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

RESULT 11  
Q9RDQ2 PRELIMINARY; PRT; 65 AA.  
ID Q9RDQ2  
AC Q9RDQ2;

Query Match 5.7%; Score 8; DB 10; Length 707;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 SAIAATVT 66  
| | | | |  
DB 41 SAIAATVT 48

DT 01-MAY-2000 (TREMELrel. 13, Created)  
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE Hypothetical protein SC02382.  
 GN SC02382 OR SC47.10.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinomycetales; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomycetes.  
 OX NCBI\_TaxID=1902;  
 RN [1]\_TaxID=1902;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.P., Harris D.E., Quail M.A., Kieser H.,  
 RA Thomson N.R., James K.D., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Harper D., Bateman A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Cronin A., Fraser A., Goble A., Larkel L., Murphy L., Oliver K., O'Neil S.,  
 RA Huang C.H., Kieser T., Lake L., Larkel L., Murphy L., Oliver K., O'Neil S.,  
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 RT coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL939112; CAB62714.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 65 AA; 6944 MW; F283FAL5A0650DCE CRC64;  
  
 Query Match 5.0%; Score 7; DB 16; Length 65;  
 Best Local Similarity 100.0%; Pred.No. 31;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 11 VAENRPT 17  
 DB 28 VAENRPT 34  
  
 RESULT 12  
 Q9BNJ8 PRELIMINARY; PRT; 91 AA.  
 AC Q9BNJ8;  
 DT 01-JUN-2001 (TREMELrel. 17, Created)  
 DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE RNA polymerase II largest subunit (Fragment).  
 OS Platydesmus sp. 'pla'.  
 OC Eukaryota; Metazoa; Arthropoda; Myriapoda; Diplopoda; Helminthomorpha;  
 OC Platydesmida; Platydesmidae; Platydesmus.  
 OX NCBI\_TaxID=118492;  
 RN [1]\_TaxID=118492;  
 RP SEQUENCE FROM N.A.  
 RA Regier J.C., Shultz J.W.;  
 RT "A phylogenetic analysis of Myriapoda (Arthropoda) using two nuclear  
 RT protein-encoding genes.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF240947; AAK11924.1; -;  
 DR InterPro; IPR007075; RNA\_pol\_Rpb1\_6.  
 DR Pfam; PF04992; RNA\_pol\_Rpb1\_6; 1.  
 FT NON\_TER 1  
 FT NON\_TER 91  
 SQ SEQUENCE 91 AA; 10270 MW; 7611D88426A06834 CRC64;  
  
 Query Match 5.0%; Score 7; DB 5; Length 91;  
 Best Local Similarity 100.0%; Pred.No. 43;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 88 RELSRKI 94  
 DB 22 RELSRKI 28  
  
 RESULT 13

Q49812 PRELIMINARY; PRT; 105 AA.  
 AC Q49812;  
 DT 01-NOV-1996 (TREMELrel. 01, Created)  
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)  
 DE B2168 C2 205.  
 OS Mycobacterium leprae.  
 OC Bacteria; Actinobacteria; Actinomycetales; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Smith D.R.;  
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Robison K.;  
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U00018; AAA17229.1; -;  
 SQ SEQUENCE 105 AA; 11450 MW; 24DDA109052F8DB0 CRC64;  
  
 Query Match 5.0%; Score 7; DB 2; Length 105;  
 Best Local Similarity 100.0%; Pred.No. 49;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 43 ALAGMCG 49  
 DB 42 ALAGMCG 48  
  
 RESULT 14  
 Q8DF92 PRELIMINARY; PRT; 133 AA.  
 AC Q8DF92;  
 DT 01-MAR-2003 (TREMELrel. 23, Created)  
 DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE Transcriptional regulator Crl-like protein.  
 OS Vibrio vulnificus.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=672;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CMCP6;  
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  
 RA Choy H.E.;  
 RT "Complete genome sequence of Vibrio vulnificus CMCP6.";  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE016798; AAC08856.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 133 AA; 15573 MW; 6B1538160953DB4F CRC64;  
  
 Query Match 5.0%; Score 7; DB 16; Length 133;  
 Best Local Similarity 100.0%; Pred.No. 60;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 83 KSPELRE 89  
 DB 51 KSPELRE 57  
  
 RESULT 15  
 Q49876 PRELIMINARY; PRT; 148 AA.  
 AC Q49876;  
 DT 01-NOV-1996 (TREMELrel. 01, Created)  
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE B229 F2 64.  
 OS Mycobacterium leprae.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1759;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Smith D.R.;  
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Robison K.;  
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U00020; AAA17318.1; -;  
 SQ SEQUENCE 148 AA; 15679 MW; 9EC6126D28CE7E86 CRC64;  
 Query Match 5.0%; Score 7; DB 2; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 66;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 STSTTN 9  
 Db 85 STSTTN 91  
 RESULT 16  
 QSVCT2  
 ID QSVCT2 PRELIMINARY; PRT; 159 AA.  
 AC QSVCT2;  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
 DE CG17244 protein (RH17470p).  
 GN CG17244.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
 RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes K., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laso P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195 (2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,  
 RA Miranda A., Mungall C.J., Nunoo J., Pacle J., Paragas V., Park S.,  
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
 RA Celnik S.;  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE003741; AAF56073.1; -;  
 DR EMBL; AY113571; AAM29576.1; -;  
 DR FlyBase; FBgn0039031; CG17244.  
 SQ SEQUENCE 159 AA; 17780 MW; 05EDC47AE278F315 CRC64;  
 Query Match 5.0%; Score 7; DB 5; Length 159;  
 Best Local Similarity 100.0%; Pred. No. 71;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 ESTSTT 8  
 Db 29 ESTSTT 35  
 RESULT 17  
 Q9BDZ0  
 ID Q9BDZ0 PRELIMINARY; PRT; 172 AA.  
 AC Q9BDZ0;  
 DT 01-JUN-2001 (TRENBLrel. 17, Created)  
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
 DE Elastin (fragment).  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;  
 OC Cercopitheidae; Macaca.  
 OX NCBI\_TaxID=9544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RA Sekhon H.S., Spindel E.R.;  
 RT "Maternal nicotine exposure up-regulates collagen and elastin gene  
 RT expression in fetal nonhuman primate lungs: potential role of alpha 7  
 RT nicotinic acetylcholine receptors.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF230927; AAK14974.1; -;  
 FT NON\_TER 172 172  
 SQ SEQUENCE 172 AA; 15224 MW; 4980EF4C5AF3330B CRC64;  
 Query Match 5.0%; Score 7; DB 6; Length 172;  
 Best Local Similarity 100.0%; Pred. No. 76;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 52 VLPGTGA 58  
 Db 105 VLPGTGA 111  
 RESULT 18  
 Q9VH39  
 ID Q9VH39 PRELIMINARY; PRT; 203 AA.  
 AC Q9VH39;  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
 DE CG11722 protein (AT14909p).



GN CG11722 OR BCDA:AT14909.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachyceta; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blake J.R., Champ M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Deicher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foglek A., Gong F., Gorrell J.H., Gu Z., Guan P., Haisler K.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champagne M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,  
 RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,  
 RA Patel S., Phuanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
 RA Celnikier S.;  
 RL Submitted (DEC-2001) to the ENBL/GenBank/DBJ databases.  
 DR EMBL; AE003685; AAF5481.1; -;  
 DR EMBL; AY070793; AAL48415.1; -;  
 DR FlyBase; FBgn003777; CG11722.  
 SQ SEQUENCE 203 AA; 23705 MW; 325F385D98DC5F99 CRC64;  
 Query Match 5.0%; Score 7; DB 5; Length 203;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 29 LLRVKSS 35  
 Db 192 LLRVKSS 198  
 |||||  
 RESULT 19  
 Q95LB5

Q95LB5 PRELIMINARY; PRT; 223 AA.  
 AC Q95LB5;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Melanocortin-4 receptor (fragment).  
 GN MC4-R.  
 OS Ovis aries (Sheep).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Caprinae; Ovis.  
 OX NCBI\_TaxID=9940;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21421436; PubMed=11530231;  
 RA Iqbal J., Pompilio S., Dumont L.M., Wu C., Mountjoy K.G., Henry B.A.,  
 RA Clarke I.J.;  
 RT "Long-term alterations in body weight do not affect the expression of  
 RT melanocortin receptor-3 and -4 mRNA in the ovine hypothalamus."  
 RL Neuroscience 105:931-940(2001).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 DR EMBL; AF345913; AAL27186.1; -;  
 DR InterPro; IPR000276; GPCR\_Rhodpsn.  
 DR Pfam; PF00001; 7tm.1; 1.  
 DR PRINTS; PR00237; GPCRHOOPS.  
 DR PROSITE; PS00237; G-PROTEIN-RECEP\_F1\_1; 1.  
 DR PROSITE; PS00262; G-PROTEIN-RECEP\_F1\_2; 1.  
 DR G-protein coupled receptor; Receptor; Transmembrane.  
 FT NON TER 1  
 FT NON TER 223  
 SQ SEQUENCE 223 AA; 24542 MW; C8EBB5FBFCAB766E CRC64;  
 Query Match 5.0%; Score 7; DB 6; Length 223;  
 Best Local Similarity 100.0%; Pred. No. 96;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 52 VLPGTGA 58  
 Db 147 VLPGTGA 153  
 |||||  
 RESULT 20  
 Q8FWQ6  
 ID Q8FWQ6 PRELIMINARY; PRT; 225 AA.  
 AC Q8FWQ6;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DE Transcriptional regulator, GntR family.  
 GN BRA0389.  
 OS Brucella suis.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Brucellaceae; Brucella.  
 OX NCBI\_TaxID=29461;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1330 / Biovar 1;  
 RX MEDLINE=22247741; PubMed=12271122;  
 RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,  
 RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,  
 RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,  
 RA Nelson W.C., Ayodeji B., Krahl M., Shetty J., Malek J., Van Aken S.E.,  
 RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,  
 RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;  
 RT "The Brucella suis genome reveals fundamental similarities between  
 RT animal and plant pathogens and symbionts."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).  
 DR EMBL; AE014538; AAN3587.1; -;  
 DR TIGR; BRA0389; -;  
 KW Complete proteome.  
 SQ SEQUENCE 225 AA; 24928 MW; 4DA74F746593C4F0 CRC64;

```

Query Match          5.0%; Score 7; DB 16; Length 225;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 MREALLR 31
    |||||
Db 57 MREALLR 63

RESULT 21
ID Q8YEM1 PRELIMINARY; PRT; 237 AA.
AC Q8YEM1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Transcriptional regulator, GNTR family.
GN BMEII0876
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20020109; PubMed=11756488;
RA Delveccio V.G., Kaparat V., Redkar R.J., Patra G., Mujar C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyripides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis";
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448 (2002).
DR EMBL; AB009721; AAL54120.1; -.
DR InterPro; IPR000524; HTH_GNTR.
DR Pfam; PF00392; gntr; 1.
DR SMART; SMO0345; HTH_GNTR; 1.
KW Complete proteome.
SQ SEQUENCE 237 AA; 26130 MW; 02655E0D6FE8BD65 CRC64;

Query Match          5.0%; Score 7; DB 16; Length 237;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 MREALLR 31
    |||||
Db 69 MREALLR 75

RESULT 22
ID Q987C4 PRELIMINARY; PRT; 243 AA.
AC Q987C4;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Transcriptional regulator.
GN MLR7110.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NAF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Ideasa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium

```

```

RT Mesorhizobium loti ";
RL DNA Res. 7:331-338 (2000).
DR EMBL; AP003011; BAB53279.1; -.
DR InterPro; IPR000524; HTH_GNTR.
DR Pfam; PF00392; gntr; 1.
DR PRINTS; PR00035; HTHGNTR.
DR SMART; SMO0345; HTH_GNTR; 1.
DR PROSITE; PS00043; HTH_GNTR_FAMILY; 1.
KW Complete proteome.
SQ SEQUENCE 243 AA; 27217 MW; 72312C98217C4B22 CRC64;

Query Match          5.0%; Score 7; DB 16; Length 243;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 PPRQST 82
    |||||
Db 76 PPRQST 82

RESULT 23
ID Q95SR4 PRELIMINARY; PRT; 255 AA.
AC Q95SR4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE GH05059p.
GN PPN OR CG1540 OR CG18436.
OS Drosophila melanogaster (Fruit fly)
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY060635; AAL28183.1; -.
DR FlyBase; FBgn0003137; Ppn.
SQ SEQUENCE 255 AA; 27727 MW; 99BF45863CFB1A63 CRC64;

Query Match          5.0%; Score 7; DB 5; Length 255;
Best Local Similarity 100.0%; Pred. No. 1.1e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 ISQESR 107
    |||||
Db 172 ISQESR 178

RESULT 24
ID Q93UL3 PRELIMINARY; PRT; 269 AA.
AC Q93UL3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Nodulation protein C (fragment).
GN NODC.
OS Rhizobium giardinii bv. phaseoli.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=142630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H251;
RX PubMed=11283294;

```

RA Laguerre G., Nour S.M., Macheret V., Sanjuan J., Drouin P.,  
 RA Amarger N.;  
 RT "Classification of rhizobia based on nodC and nifH gene analysis  
 RT reveals a close phylogenetic relationship among Phaseolus vulgaris  
 RT symbionts.";  
 RL Microbiology 147:981-993(2001).  
 DR EMBL: AF217264; AAK39959.1; -.  
 DR InterPro; IPR001064; Crystallin.  
 DR InterPro; IPR004835; Fungi chitin syn.  
 DR Pfam; PF03142; Chitin synth 2; 1.  
 DR Pfam; PF00535; Glycos\_transf\_2; 1.  
 DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
 FT NON\_TER 1  
 FT NON\_TER 269  
 SQ SEQUENCE 269 AA; 29825 MW; 02BBA4734884D175 CRC64;

Query Match 5.0%; Score 7; DB 2; Length 269;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 59 SAIATV 65  
 Db 148 SAIATV 154  
 |||||

RESULT 25  
 Q93UK8 PRELIMINARY; PRT; 270 AA.  
 AC Q93UK8  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Nodulation protein C (fragment).  
 GN NODC.

OS Sinorhizobium sp. GR-06.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=147701;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GR-06;  
 RX PubMed=11283294;

RA Laguerre G., Nour S.M., Macheret V., Sanjuan J., Drouin P.,  
 RA Amarger N.;  
 RT "Classification of rhizobia based on nodC and nifH gene analysis  
 RT reveals a close phylogenetic relationship among Phaseolus vulgaris  
 RT symbionts.";  
 RL Microbiology 147:981-993(2001).  
 DR EMBL: AF217269; AAK39964.1; -.  
 DR InterPro; IPR001064; Crystallin.  
 DR InterPro; IPR004835; Fungi chitin syn.  
 DR Pfam; PF03142; Chitin synth 2; 1.  
 DR Pfam; PF00535; Glycos\_transf\_2; 1.  
 DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
 FT NON\_TER 1  
 FT NON\_TER 270  
 SQ SEQUENCE 270 AA; 30089 MW; 5841B33D6EEDD131 CRC64;

Query Match 5.0%; Score 7; DB 2; Length 270;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 59 SAIATV 65  
 Db 154 SAIATV 160  
 |||||

RESULT 26  
 Q8RNL4 PRELIMINARY; PRT; 273 AA.  
 ID Q8RNL4  
 AC Q8RNL4

DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE NodC (fragment).  
 OS Sinorhizobium sp. 16b1.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=189051;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=16b1;  
 RA Mhamdi R., Laguerre G., Aouani M.E., Mars M., Amarger N.;  
 RT "Different species and symbiotic genotypes of field rhizobia can  
 RT nodulate Phaseolus vulgaris in Tunisian soils.";  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF481764; AAL88670.1; -.  
 DR InterPro; IPR001064; Crystallin.  
 DR InterPro; IPR004835; Fungi chitin syn.  
 DR Pfam; PF03142; Chitin synth 2; 1.  
 DR Pfam; PF00535; Glycos\_transf\_2; 1.  
 DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
 FT NON\_TER 1  
 FT NON\_TER 273  
 SQ SEQUENCE 273 AA; 30557 MW; 2FFA2B2343E71325 CRC64;

Query Match 5.0%; Score 7; DB 2; Length 273;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 SAIATV 65  
 Db 160 SAIATV 166  
 |||||

RESULT 27  
 Q93UL5 PRELIMINARY; PRT; 281 AA.  
 AC Q93UL5  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Nodulation protein C (fragment).  
 GN NODC.

OS Rhizobium etli bv. phaseoli.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium.  
 OX NCBI\_TaxID=147700;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Vikingi;  
 RX PubMed=11283294;

RA Laguerre G., Nour S.M., Macheret V., Sanjuan J., Drouin P.,  
 RA Amarger N.;  
 RT "Classification of rhizobia based on nodC and nifH gene analysis  
 RT reveals a close phylogenetic relationship among Phaseolus vulgaris  
 RT symbionts.";  
 RL Microbiology 147:981-993(2001).  
 DR EMBL: AF217262; AAK39957.1; -.  
 DR InterPro; IPR001064; Crystallin.  
 DR InterPro; IPR004835; Fungi chitin syn.  
 DR Pfam; PF03142; Chitin synth 2; 1.  
 DR Pfam; PF00535; Glycos\_transf\_2; 1.  
 DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
 FT NON\_TER 1  
 FT NON\_TER 281  
 SQ SEQUENCE 281 AA; 31326 MW; D46A13EA185121C5 CRC64;

Query Match 5.0%; Score 7; DB 2; Length 281;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 SAIAATV 65  
Db 163 SAIAATV 169

RESULT 28  
Q93UL4 PRELIMINARY; PRT; 282 AA.  
AC Q93UL4  
DT 01-DEC-2001 (TReMBLrel. 19, Created)  
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
DE Nodulation protein C (Fragment).  
GN NODC.  
OS Rhizobium leguminosarum (biovar phaseoli).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=385;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H132;  
RX PubMed=11283294;  
RA Laguerre G., Nour S.M., Macheret V., Sanjuan J., Drouin P.,  
RA Amarger N.;  
RT "Classification of rhizobia based on nodC and nifH gene analysis  
RT reveals a close phylogenetic relationship among Phaseolus vulgaris  
RT symbionts.";  
RL Microbiology 147:981-993(2001).  
RL EMBL; AF217263; AAK3958.1; -;  
DR InterPro; IPR001064; Crystallin.  
DR InterPro; IPR004835; Fungi chitin syn.  
DR InterPro; IPR001173; Glyco\_transf\_2.  
DR Pfam; PF03142; Chitin synthase 2; 1.  
DR Pfam; PF00535; Glycos\_transf\_2; 1.  
DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
FT NON\_TER 1  
FT NON\_TER 282  
SQ SEQUENCE 282 AA; 31401 MW; A13C34694261A9DB CRC64;

Query Match 5.0%; Score 7; DB 2; Length 282;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 SAIAATV 65  
Db 164 SAIAATV 170

RESULT 29  
Q8DID3 PRELIMINARY; PRT; 283 AA.  
AC Q8DID3  
DT 01-MAR-2003 (TReMBLrel. 23, Created)  
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
DE Tlr1657 protein.  
GN TLR1657.  
OS Synecococcus elongatus (Thermosynechococcus elongatus).  
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.  
OX NCBI\_TaxID=32046;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BP-1;  
RX MEDLINE=22255444; PubMed=12240834;  
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,  
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,  
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,  
RA Shimpou S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the thermophilic cyanobacterium  
RT Thermosynechococcus elongatus BP-1";  
RL DNA Res. 9:123-130(2002).  
RL EMBL; AF005374; BAC09209.1; -;  
KW Complete proteome.

SQ SEQUENCE 283 AA; 30758 MW; D2CEF14B10653CFF CRC64;  
Query Match 5.0%; Score 7; DB 16; Length 283;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 EALLRVK 33  
Db 266 EALLRVK 272

RESULT 30  
Q8U2B4 PRELIMINARY; PRT; 283 AA.  
ID Q8U2B4  
AC Q8U2B4  
DT 01-JUN-2002 (TReMBLrel. 21, Created)  
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
DE Heme biosynthesis protein.  
GN PF0925.  
OS Pyrococcus furiosus.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Pyrococcus.  
OX NCBI\_TaxID=2261;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;  
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;  
RT "The complete sequence of the Pyrococcus furiosus genome.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE010206; AAL81049.1; -;  
KW Complete proteome.  
SQ SEQUENCE 283 AA; 32336 MW; 779EBE38D867B066 CRC64;

Query Match 5.0%; Score 7; DB 17; Length 283;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 ELSRKIR 95  
Db 273 ELSRKIR 279

RESULT 31  
Q93UK9 PRELIMINARY; PRT; 286 AA.  
ID Q93UK9  
AC Q93UK9  
DT 01-DEC-2001 (TReMBLrel. 19, Created)  
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
DE Nodulation protein C (Fragment).  
GN NODC.  
OS Rhizobium etli bv. phaseoli.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=147700;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CFN42;  
RX PubMed=11283294;  
RA Laguerre G., Nour S.M., Macheret V., Sanjuan J., Drouin P.,  
RA Amarger N.;  
RT "Classification of rhizobia based on nodC and nifH gene analysis  
RT reveals a close phylogenetic relationship among Phaseolus vulgaris  
RT symbionts.";  
RL Microbiology 147:981-993(2001).  
RL EMBL; AF217268; AAK39563.1; -;  
DR InterPro; IPR001064; Crystallin.  
DR InterPro; IPR004835; Fungi chitin syn.  
DR InterPro; IPR001173; Glyco\_transf\_2.  
DR Pfam; PF03142; Chitin synthase 2; 1.  
DR Pfam; PF00535; Glycos\_transf\_2; 1.  
DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.

FT NON\_TER 1 1  
FT NON\_TER 286 286  
SQ SEQUENCE 286 AA; 31784 MW; 90E168AB090F7B5B CRC64;  
Query Match 5.0%; Score 7; DB 2; Length 286;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 59 SAIAATV 65  
DB 161 SAIAATV 167  
RESULT 32  
QY 59 SAIAATV 65  
DB 161 SAIAATV 167  
ID Q93UL2 PRELIMINARY; PRT; 290 AA.  
AC Q93UL2;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Nodulation protein C (Fragment).  
GN NODC.  
OS Rhizobium gallicum bv. phaseoli.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=142628;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Phd12;  
RX PubMed=11283294;  
RA Laguerre G., Nour S.M., Macheret V., Sanjuan J., Drouin P.,  
Amarger N.;  
RT "Classification of rhizobia based on nodC and nifH gene analysis  
RT reveals a close phylogenetic relationship among Phaseolus vulgaris  
RT symbionts.";  
RL Microbiology 147:981-993(2001).  
DR EMBL; AF217265; AAK39960.1; -;  
DR InterPro; IPR001064; Crystallin.  
DR InterPro; IPR004835; Fungi\_chitin\_syn.  
DR InterPro; IPR001173; Glyco\_trans.2.  
DR Pfam; PF03142; Chitin\_synth.2; 1.  
DR Pfam; PF00535; Glycos\_transf.2; 1.  
DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
FT NON\_TER 1 1  
FT NON\_TER 290 290  
SQ SEQUENCE 290 AA; 32213 MW; 7F9BD793B399F584 CRC64;  
Query Match 5.0%; Score 7; DB 2; Length 290;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 59 SAIAATV 65  
DB 165 SAIAATV 171  
RESULT 33  
ID O56780 PRELIMINARY; PRT; 297 AA.  
AC O56780;  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Phosphoprotein.  
GN P.  
OS European bat lyssavirus 2.  
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
OC Rhabdoviridae; Lyssavirus.  
OX NCBI\_TaxID=57483;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=V286;  
RA Nadin-Davis S.A., Abdel-Malik M., Huang W., Armstrong J.,

RA Wandeler A.I.;  
RT "Genetic characterisation of the Lyssavirus P gene locus.";  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF049121; AAC04591.1; -;  
DR InterPro; IPR004259; PP\_M1.  
DR Pfam; PF03012; PP\_M1; 1.  
SQ SEQUENCE 297 AA; 33287 MW; 52D56CSEA902650 CRC64;  
Query Match 5.0%; Score 7; DB 12; Length 297;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 99 KTISOES 105  
DB 172 KTISOES 178  
RESULT 34  
ID Q9CIK6 PRELIMINARY; PRT; 308 AA.  
AC Q9CIK6;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Snz-type pyridoxine vitamin B6 biosynthetic protein SNZ1.  
GN PDX-1.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=74-OR23-1A;  
RX MEDLINE=21135672; PubMed=11238395;  
RA Baan L.E., Dvorachek W.H. Jr., Braun E.L., Errett A., Saenz G.S.,  
Giles M.D., Werner-Washburne M., Nelson M.A., Natvig D.O.;  
RT "Analysis of the pdx-1 (snz-1/sno-1) region of the Neurospora crassa  
RT genome. Correlation of pyridoxine-requiring phenotypes with mutations  
RT in two structural genes.";  
RL Genetics 157:1067-1075(2001).  
DR EMBL; AF309689; AAK07850.1; -;  
DR InterPro; IPR003009; FMN\_enzyme.  
DR InterPro; IPR001852; Snzlp/Sori.  
DR Pfam; PF01680; SOR\_SNZ; 1.  
DR ProDom; PD004958; Snzlp/Sori; 1.  
DR TIGRFAMs; TIGR00343; TIGR00343; 1.  
DR PROSITE; PS01235; UPF0019; 1.  
SQ SEQUENCE 308 AA; 32355 MW; 47D73BB4CAD284D7 CRC64;  
Query Match 5.0%; Score 7; DB 3; Length 308;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3 STSTTN 9  
DB 3 STSTTN 9  
RESULT 35  
ID Q93UK6 PRELIMINARY; PRT; 311 AA.  
AC Q93UK6;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Nodulation protein C (Fragment).  
GN NODC.  
OS Rhizobium leguminosarum (biovar trifolii).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=386;  
RN [1]  
RP SEQUENCE FROM N.A.

```

RC STRAIN=USDA2071;
RX PubMed=11283294;
RA Laguerre G., Nour S.M., Macheret V., Sanjuan J., Drouin P.,
RA Amarger N.;
RT "Classification of rhizobia based on nodC and nifH gene analysis
RT reveals a close phylogenetic relationship among Phaseolus vulgaris
RT symbionts.";
RL Microbiology 147:981-993(2001).
DR EMBL; AF217271; AAK39966.1; -.
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR004835; Fungi chitin syn.
DR InterPro; IPR001173; Glyco-trans_2.
DR Pfam; PF03142; Chitin_synth_2; 1.
DR Pfam; PF00335; Glycos_transf_2; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
FT NON_TER
FT 1
FT 311
SQ SEQUENCE 311 AA; 34360 MW; B3C03D399D59081B CRC64;

Query Match 5.0%; Score 7; DB 2; Length 311;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 SAIAATV 65
DB 176 SAIAATV 182

RESULT 36
Q8XGX9 PRELIMINARY; PRT; 323 AA.
ID Q8XGX9
AC Q8XGX9
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Preprotein translocase, IISF family, membrane subunit (Protein-export
DE membrane protein SecY).
GN SECY OR STM0408 OR STY0446.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking I., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RX EMBL; AE008714; AAL19362.1; -.
DR EMBL; AL627266; CAD08864.1; -.
DR InterPro; IPR003395; SecD_SecF.

InterPro; IPR005665; SecF.
Pfam; PF02355; SecD_SecF; 1.
TIGRFAMs; TIGR00916; 2A0604801; 1.
TIGRFAMs; TIGR00966; 3a0501807; 1.
KW Complete proteome.
SQ SEQUENCE 323 AA; 35376 MW; 10D06394DF60C886 CRC64;

Query Match 5.0%; Score 7; DB 16; Length 323;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 DVMREAL 29
DB 69 DVMREAL 75

RESULT 37
Q9I4V9 PRELIMINARY; PRT; 330 AA.
ID Q9I4V9
AC Q9I4V9
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Probable glycosyl transferase.
GN PA1014.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Huynaghe W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Brody L.L., Coulter S.N., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
RX EMBL; AE004534; AAG04403.1; -.
DR HSSP; P39621; IQGQ.
DR InterPro; IPR001173; Glyco_trans_2.
DR Pfam; PF00535; Glycos_transf_2; 1.
KW Transferase; Complete_proteome.
SQ SEQUENCE 330 AA; 37355 MW; 77FB8654AEE29F86 CRC64;

Query Match 5.0%; Score 7; DB 16; Length 330;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 ERLAMLR 42
DB 79 ERLAMLR 85

RESULT 38
Q9AD29 PRELIMINARY; PRT; 331 AA.
ID Q9AD29
AC Q9AD29
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein SCPI.121.
GN SCPI.121.
OS Streptomyces coelicolor.
OG Plasmid SCPI.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]

```

RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Brown S.P., Murphy L.D., Harris D.;  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE=98241550; PubMed=9573173;  
RA Redenbach M., Ikeda K., Yamasaki M., Kinashi H.;  
RT "Cloning and physical mapping of the EcoRI fragments of the giant  
RT 1 linear plasmid SCPI";  
RL J. Bacteriol. 180:2796-2799 (1998).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE=21996410; PubMed=12000953;  
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
RT coelicolor A3(2)";  
RL Nature 417:141-147 (2002).  
DR EMBL; AL590463; CAC36642.1; -;  
DR InterPro; IPR006209; EGF like.  
DR PROSITE; PS01186; EGF 2; 1.  
KW Hypothetical protein; Plasmid; Complete proteome.  
SQ SEQUENCE 331 AA; 36661 MW; DF1384E21D3434DC CRC64;  
  
Query Match 5.0%; Score 7; DB 16; Length 331;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 49 GHRVLP 55  
DB 97 GHRVLP 103  
  
RESULT 39  
Q8HXX3 PRELIMINARY; PRT; 332 AA.  
AC Q8HXX3;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Melanocortin-4 receptor.  
GN MC4R.  
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecinae; Macaca.  
OX NCBI\_TaxID=9541;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain cerebellum cortex;  
RA Kusuda J., Osada N., Hida M., Sugano S., Hashimoto K.;  
RT "Isolation and characterization of cDNA for macaque neurological  
RT disease genes";  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB083317; BAC20596.1; -;  
KW Receptor.  
SQ SEQUENCE 332 AA; 37001 MW; 400E0CEB110FD8C7 CRC64;

Query Match 5.0%; Score 7; DB 6; Length 332;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 52 VLPCTGA 58  
DB 228 VLPCTGA 234  
  
RESULT 40  
Q97YG7 PRELIMINARY; PRT; 332 AA.  
AC Q97YG7;  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Endoglucanase precursor (EC 3.2.1.4).  
GN SSO1354.  
OS Sulfolobus solfataricus.  
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
OC Sulfolobus.  
OX NCBI\_TaxID=2287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=ATCC 35092 / DSM 1617 / P2;  
RX MEDLINE=21332296; PubMed=1142726;  
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
RA Aways M.J., Chan-Weiner C.C.-F., Clausen I.G., Curtis B.A.,  
RA De Moors A., Erauso G., Fietcher C., Gordon P.M.K.,  
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.,  
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840 (2001).  
DR EMBL; AE006749; AAK41590.1; -;  
DR InterPro; IPR002594; Glyco\_hydro\_12.  
DR Pfam; PF01670; Glyco\_hydro\_12; 1.  
DR ProDom; PD004316; Glyco\_hydro\_12; 1.  
KW Hydrolase; Glycosidase; Complete proteome.  
SQ SEQUENCE 332 AA; 37295 MW; F096165AE4926B9E CRC64;  
  
Query Match 5.0%; Score 7; DB 17; Length 332;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 3 STSTTN 9  
DB 45 STSTTN 51  
  
Search completed: November 14, 2003, 10:52:09  
Job time : 65 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 17, 2003, 12:05:08 ; Search time 2089 seconds  
(without alignments)  
2761.253 Million cell updates/sec

Title: US-10-087-573-2

Perfect score: 708

Sequence: 1 MESTSTTTNFVAENPTFGE.....RAEYFRHLRLSLKSGQGNRLI 141

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+pn.model -DEV=xlh

-Q/cgn2\_1/USPTO.spool/US10087573/runat\_14112003\_103610\_20673/app.query.fasta\_1.327

-DB=GenEmbl -QFMT=fastap -SUFFIX=rg -MINMATCH=0.1 -LOOFC=0 -LOOPEXT=0

-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45

-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL

-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US10087573 @cgn\_1\_1.2372 @runat\_14112003\_103610\_20673 -RCPU=6 -ICPU=3

-NO WAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG

-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	708	100.0	1115	3	BCA494862
2	708	100.0	1135	6	AX530469
3	680	96.0	1134	6	AX530471
4	95.5	13.5	10463	1	AE012139
5	92.5	13.1	4656	9	HSM803142
6	92.5	13.1	4839	6	AX430339
7	92.5	13.1	5973	1	AB024056
8	92.5	13.1	6573	1	AB027763
9	92.5	13.1	10123	6	AX367101
10	92.5	13.1	10759	6	BD084093
11	92.5	13.1	10759	9	AB053446
12	92.5	13.1	192239	9	AC009796
13	92.5	13.1	208430	2	AC027641
14	92.5	13.1	211735	9	AC091564
15	92	13.0	90	6	AX530473
16	92	13.0	11060	1	AE004618
17	91.5	12.9	1386	10	AF416730
18	91.5	12.9	1607	10	BC002136
19	91.5	12.9	1618	10	AF495469
20	91.5	12.9	1633	10	AF143955
21	90	12.7	1563	9	HSCORONIN
22	90	12.7	1589	9	HSU34690
23	90	12.7	1639	9	AF495470
24	90	12.7	1657	6	E11332
25	90	12.7	1659	6	AX237691
26	90	12.7	1659	9	HUMP57B
27	90	12.7	208443	2	AC120207
28	90	12.7	210614	1	AB088224
29	90	12.7	217346	9	AC012363
30	89.5	12.6	1227	1	TFEMADF
31	89	12.6	96236	3	AC012492
32	88.5	12.5	176562	9	AC084864
33	88.5	12.5	198677	1	AE001863
34	88	12.4	846	6	BD180073
35	88	12.4	148797	2	AP005426
36	88	12.4	156242	2	AP005429
37	88	12.4	162382	2	AP005180
38	88	12.4	283100	1	SC0939110
39	88	12.4	300813	2	AC125900
40	88	12.4	308821	2	AC118450
41	88	12.4	348430	2	AC130157
42	87.5	12.4	1386	10	AF047388
43	87.5	12.4	1605	10	BC053398
44	87.5	12.4	1616	10	AF495468
45	87.5	12.4	2115	5	BC047956

# ALIGNMENTS

RESULT 1



US-10-087-573-2 (1-141) x BCA494862 (1-1115)	
Qy	1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20
Db	61 ATGGAGTCGACATCAACACGACCAACTTTGTTGCCGAGAACCGTCCACCTTTGGTGAG 120
Qy	21 ThrPheAspValMetArgGluAlaLeuLeuArgValValSerSerGluArgLeuAlaMet 40
Db	121 ACCTTTGATGTGATGAGGGAAGCTTTGCTTCGTGTAAAGTCCTCTGAACGCTTGGCAATG 180
Qy	41 LeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuLeuProGlyThrGlyAlaSerAla 60
Db	181 CTCAGAGCGCTTCAGAGATGTCGGTCACCGCGTCTCTCTGGCACTGGTGCTTCTGCG 240
Qy	61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProArgProGln 80
Db	241 ATAGCGCAACGGTAACCCCAAGGGGGCTTCGATGAAGCTTAAACACACCGCTCGCAG 300
Qy	81 SerThrLysSerProGluLeuArgGluLeuSerArgLysAlaLeuArgGluMetAsnLysThr 100
Db	301 TCAACGAAGTCTCCGAGCTCAGGAGCTGTACGGAAGATTCGCGAATGAATAAGACT 360
Qy	101 IleSerGlnGluSerAlaArgValAsnHisArgLeuLeuProGluGlyHisProLeuLeuGlu 120
Db	361 ATAAGTCAGGAATCAGCTCGGTAAACACCGGTTCCCGAAGCCACCTCTCTTAGAG 420
Qy	121 LysArgAlaGluTyrrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140
Db	421 AACCGGCGAGATATTTTCGTCACTTAGATCTCTTTAAGAGCCAGAGAGTCAATAGACTC 480
Qy	141 Ile 141
Db	481 ATC 483
RESULT 2	
AX530469	AX530469 1135 bp DNA linear PAT 22-NOV-2002
LOCUS	Sequence 1 from Patent EP1238983.
DEFINITION	
ACCESSION	AX530469
VERSION	AX530469.1 GI:25252327
KEYWORDS	
SOURCE	Babesia canis
ORGANISM	Babesia canis
	Eukaryota; Alveolata; Apicomplexa; Piropalmsida; Babesiidae;
REFERENCE	
AUTHORS	Schettlers, T.P., Carcy, B.P., Drakulovski, P.R. and Gorenflot, A.F.
TITLE	Babesia canis vaccine
JOURNAL	Patent: EP 1238983-A 11-SEP-2002;
	Akzo Nobel N.V. (NL)
FEATURES	Location/Qualifiers
source	1..1135
	/organism="Babesia canis"
	/mol_type="genomic DNA"
	/db_xref="taxon:5867"
	75..500
	/note="unnamed protein product"
	/codon_start=1
	/protein_id="CAD57205.1"
	/db_xref="GI:25252328"
	/translation="WESTSTITNFVAENPPTFGTFDVMREALLRVKYSERLAMLRLAL
	AGMCHRLVPGTGAISAIAATVTFKASMKLKPQPQSTKSPRLSLRSKIREMNTKIS
	QESARVNHRLPEGHLEKRAEYFRHLRLSKSQGVNRLI"
BASE COUNT	308 a 286 c 289 g 252 t

US-10-087-573-2 (1-141) x AX530469 (1-1135)		Percent Similarity: 99.29%	Conservative: 0
		Best Local Similarity: 99.29%	Mismatches: 1
		Query Match: 96.05%	Indels: 1
		DB: 6	Gaps: 0
US-10-087-573-2 (1-141) x AX530471 (1-1134)			
Qy	1	MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu	20
Db	75	ATGGAGTCGACATCAACACGACCACTTTGTTGCCGAGAACCGTCCACCTTTGGTGAG	134
Qy	21	ThrPheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMet	40
Db	135	ACGTTTGTATGTGATGAGGGAAGCTTTGCTTCGTGTAAAGTCTCTGAACGCTTGGCAATG	194
Qy	41	LeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla	60
Db	195	CTCAGAGCGCTTCGAGGAATGTCCGGTCAACCGGCTCTTCTGGCACTGGTCTTCTGCG	254
Qy	61	IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgProGln	80
Db	255	ATAGCGGCAACGGTAAACCCCAAGGGGGCTTCGATGAAGCTTAAACACCGCGTCCGCG	314
Qy	81	SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThr	100
Db	315	TCAACGAAGTCTCCGGAGCTCAGGAGCTGTACGGAAGATTCGCGAAATGAATAGACT	374
Qy	101	IleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGlu	120
Db	375	ATAAGTCAGGAATCAGCTCGGTAAACCCCGGTTCGCGAAGCCACCTCTCTTAGAG	434
Qy	121	LysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu	140
Db	435	AAGCGGGCAGATATTTCGTCACTTAGATCTCTTAGAGCCAGAGTCAATAGACTC	494
Qy	141	Ile 141	
Db	494	ATC 496	
RESULT 4			
AX530471			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
JOURNAL			
FEATURES			
Source			
CDS			
BASE COUNT			
ORIGIN			
Alignment Scores:			
Pred. No.:			
Score:			
TITLE			
MEDLINE			
PUBMED			
Xanthomonas campestris pv. campestris str. ATCC 33913			
Xanthomonas campestris pv. campestris str. ATCC 33913			
Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;			
Xanthomonadaceae; Xanthomonas.			
1 (bases 1 to 10463)			
da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,			
Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida			
Jr., N.P., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,			
Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardozo, J.,			
Chambergo, F., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L.,			
Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,			
Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,			
Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite			
Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,			
Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,			
Menes, C.F.M., Miyaki, C.T., Moon, D.H., Moreira, L.M., Novo, M.T.M.,			
Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,			
Rosa, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,			
Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos			
Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and			
Kitajima, J.P.			
Comparison of the genomes of two Xanthomonas pathogens with			
differing host specificities			
Nature 417 (6887), 459-463 (2002)			
22022145			
12024217			

REFERENCE  
AUTHORS

2 (bases 1 to 10463)  
da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,  
Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida  
Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, W.C.,  
Camargo, L.B.A., Camarotte, G., Cannavan, F., Cardozo, J.,  
Chamargo, F., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L.,  
Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,  
Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,  
Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite  
Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,  
Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,  
Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,  
Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,  
Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,  
Taktá, M.A., Tamura, R.E., Teixeira, E.C., Tezari, R.I.D., Trindade dos  
Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and  
Kitajima, J.P.

TITLE  
JOURNAL

Direct Submission  
Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de  
Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,  
Brazil

FEATURES  
source

Location/Qualifiers  
1. .10463  
/organism="Xanthomonas campestris pv. campestris str. ATCC  
33913"  
/mol\_type="genomic DNA"  
/strain="ATCC 33913"  
/db\_xref="ATCC:33913"  
/db\_xref="taxon:190485"  
/note="pathovar: campestris"  
complement(457. .1215)  
/genes="XCC0416"  
complement(457. .1215)  
/genes="XCC0416"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=11  
/product="3-oxoacyl-[ACP] reductase"  
/protein\_id="AA039734.1"  
/db\_xref="GI:21111399"  
/translation="MSTPPQPVRLVIGSGIGLATAQFAQVQSGAQSICARNPEG  
LANAAQLAAGAPVHTFACDLADAQERYVHAQAQFGLGVVNVNAGSGHND  
ESWQGLVDLMAVRCNAAPVYLRQSGNAVILNMISSNGORPPFRAIAYSTAKAL  
NYTTTLAELARIRVNAIPAGSIEFGLWEQRSDPALYARIRDSIPFGFGQ  
VQHIADAULFLASPOARWITGQVLAVDQGSILGV"  
1518. .2954  
/genes="XCC0417"  
1518. .2954  
/genes="XCC0417"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA039735.1"  
/db\_xref="GI:21111400"  
/translation="VQVFERELHMTHTVIGAGCGSVLVRALAQGADAQVRLTVG  
VETFGSGIAYGAARPEHLNVRKADLQDAQAPAFADTLHGESGLLEPLRAYG  
DYRLSELDITGAQAQASIVRLSQEAVAVERRGRFVFLANGDAFQTDQVVLAVGALQ  
PQALAGIGPLSPHYRIGWPQDALSPPDDVLI VGTGLTMDVVALSLHAGHR  
GRLIATSRGLAPQALHQPQGALELPHLQRAIKDADRLALRGVROLVVVDWRR  
VVDALRPHLQPIWQLELSORAFRLHLRYPMEVHARVAPCAADQLAQDQSCLOV  
VAALLRRAVPDGVAVIRPFGGADCTRRFDVAVRATGLDITDRTSDLLAGMRE  
AGLLRADPLGLCVDTDAQLRVDPDQGVPGLYCGVPLLRGRYWEITAVPELRVATRA  
LAERLLQATPALPQVQPDVTPREVNRP"  
3277. .3708  
/gene="virK"  
/note="synonym: XCC0418"  
3277. .3708  
/gene="virK"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

gene

gene  
CDS

/protein id="AAM39740.1"  
/db\_xref="GI:21111405"  
/translation="MSEGRNLSALAVRERSITFLVLISLAGLVAFLKLRADPPA  
FTVKYMTIITAPGATPQMEQVAKLEKMQELRWDRITETIRPGLATTTLLD  
STPGEVQGFQARKKYNIGQPERIFVESHDLATLGVSPQVFAALNAQNA  
ADAETMRQLHVPKVKYNIIGQPERIFVESHDLATLGVSPQVFAALNAQNA  
LTAAGSVTRGDFIIRDLGDLNLOKIRDTPLVQGRSLKLSDIATVKRVEDPSTF  
MIRSGEPALLGIIMRDGNDLGLKDAEVAENAEPLGMQLSKVTQAVNIDA  
SVGEFTKFFVALLVVLVCFVSMGRVGVVAAAVPLTAAVFMVLATCKNDRIT  
LQSLALGLLVDLAIIEEMVVMKEGGRVAAAVAMSHTAAPMLSGTLVAVGF  
MNGFAASTAGEYTSNPKIVGIALIVSAVAVVTPYLVKMLPDMKKEGGAALY  
NTPYRNRQALGRVIARKLVAGSVGLFVAVVGMGIVKQFFPSIDREVLVEVQ  
LPGYSITAAAKSAMLSKQEKIIVTAVVGMGIVKQFFPSIDREVLVEVQ  
VTDQDHQMDKLRALAGLAPEARVATQTFGYSKFPVAVVSGPDPSTVLR  
GIAQVQMDQSLRNTVNTDMVGTPTLHSLDQDLQAVGLTSAVAQLOFLIS  
GVPIILVREDIRSVQVARSAGDRLDPAIADFTLAGNGORVPLSOVGKVDYRME  
PVMRRDRVPTITCGVDVDTLOPDVSAITROLOPIIDKLPSGYOIKKAGSIEES  
KATAAMLPLPPLMLAATLIIILQVRSISAMVMFELTSLGLIGVPTLILFQGFPGI  
NALVGLIALSGILMENTLIILOVRSISAMVMFELTSLGLIGVPTLILFQGFPGI  
AFIPHTSVFNGTLATLUGTGLATLVTLLVFLPAMYSIWFKIRPDGKGNAPDIAR  
FA"

BASE COUNT 1668 a 3424 c 3575 g 1796 t

Alignment Scores:  
Pred. No.: 305 Length: 10463  
Score: 95.50 Matches: 38  
Percent Similarity: 38.4% Conservative: 17  
Best Local Similarity: 26.5% Mismatches: 46  
Query Match: 13.4% Indels: 42  
DB: 1 Gaps: 4

US-10-087-573-2 (1-141) x AE012139 (1-10463)

QY 26 ArgGluAlaLeuArgValLysSerGluArgLeuAlaMetLeuArgAlaLeuAla 45  
Db 2018 CGGCGCGGTTTGAGCGTGCACCCGCGCTACATCGGCTG-----GCCGTGGCA 2065  
QY 46 GlyMetCysGlyHisArgValLeuPro-----54  
Db 2066 GGGCGATGCGTTGTGCGGTTTGTCCCGACGATGACGTGCTGATCGTGTGACCGGCT 2125  
QY 55 -----GlyThrGlyAlaSer-----59  
Db 2126 GACCATGTGAGCGTGGCGCTGAGCTTCCACGACGTGCGCCACCGCGCGCTGCTGGC 2185  
QY 60 -----AlaIleAlaAlaThrValThrPro-LysGlyAlaSerMetLy 73  
Db 2186 GATCTCGCGCGCGCTGGCGACCGCGCATCTACCGACGCGAGTGGCGGCTGGGA 2245  
QY 73 sLeuLysProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLy 93  
Db 2246 GCTGCGCGCGCATCTGCAACGACGCGCATCAAGGATGCGCGCGCTGTTGCGTGG 2305  
QY 93 sIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeuPr 113  
Db 2306 CGTGGCCCACTGAGCGTGGTGGAGCATGGCGGCGGCGTGGTGGATGCGTGGCGCC 2365  
QY 113 oGluGlyHisProLeu-----LeuGluLysArgAlaGluThrPheArgHi 128  
Db 2366 GCATCTGAGCGCGCTCTGGCAACGCTTTGGAGCTAAGCAACGCGCGCGCTTCTACGCCA 2425  
QY 128 sLeuArg 130  
Db 2426 TTTGCGC 2432

RESULT 5  
HSM803142/c  
LOCUS  
DEFINITION Homo sapiens mRNA; cDNA DKFZp547P1613 (from clone DKFZp547P1613).  
ACCESSION AL331828  
VERSION AL331828.1 GI:21732302  
KEYWORDS  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 4656)  
AUTHORS Koehrer, X., Beyer, A., Mewes, H.W., Weil, B. and Wiemann, S.  
TITLE Direct Submission  
JOURNAL Submitted (09-JUL-2002) 1. D-85764 Neuherberg, GERMANY  
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
sequenced by BMFZ (Biomedical Research Center at the Charite,  
Berlin/Germany) within the cDNA sequencing consortium of the German  
Genome Project.  
This clone (DKFZp547P1613) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clonesrzd.de Further  
information about the clone and the sequencing project is available  
at http://mips.gsf.de/proj/cDNA/.  
FEATURES  
source  
1. 4656  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/clone="DKFZp547P1613"  
/tissue\_type="brain"  
/clone\_lib="547 (synonym: hfbrl). Vector pSport1; host  
DH10B; sites NotI + SalI"  
/dev\_stage="fetal"  
1. 4656  
/gene="DKFZp547P1613"  
3792..4412  
/gene="DKFZp547P1613"  
/notes="similarity to cadherins"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAD38541.1"  
/db\_xref="GI:21732303"  
/translation="MKLGHPQLPVSTHPPSLPWTQBPSPCLPSPQQTQLQPGPSH  
QLTAPSAVKAPCPQPCPASHPLCLLMLAHLPLHGLWPRVPQPHSAQSLWS  
HLNTRSCSGCGPWAPTDAHSVFNAGPTLSLPWAASDYDLWLRGHGHPALHPPLP  
SPLRVLGPKQKCSLTCDQVQCKEYEGGSSPGFSSVRDPCPAPAP"  
4558..4563  
/gene="DKFZp547P1613"  
4574  
/gene="DKFZp547P1613"  
BASE COUNT 953 a 1467 c 1307 g 929 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 215 Length: 4656  
Score: 92.50 Matches: 38  
Percent Similarity: 46.58% Conservative: 30  
Best Local Similarity: 26.03% Mismatches: 52  
Query Match: 13.06% Indels: 26  
DB: 9 Gaps: 6  
US-10-087-573-2 (1-141) x HSM803142 (1-4656)  
QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
Db 3084 GCAAGCACCACCACTCCCAAGGAGGTCGCCCGCCCTACTAATAGC-----3034  
QY 23 AspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
Db 3033 -----AGGTGAGGTGAGTGCAGGCCAGGCCAGTGGCGGTGG 2998  
QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeu-----ProGlyThrGlyAlaSerAla 60  
Db 2997 GTGATATCCACCGTCCAGGACCTGTGGCACTCCCGGAACACGAGGAGGCCCCCGTGT 2938  
QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProArgProGln 80  
Db 2937 ATCACTCCAGCTCAGCTCCGCTGCTTCCCGCGGTCGCGGCCCCAGAG 2878

```
QY 81 ---SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArg----- 95
Db 2877 GTGGCTGTTCCGCTGCTGCTGCCGACGTGTCACCCGAGGTACAGGCTCTGTAGTC 2818
QY 96 -----GluMetAsnLysThrIleSerGlnGluSerAlaArgVal-AsnHisAr 111
Db 2817 TGTTTAATACCAAAATAGGGGAAGAGTGGCAAGGGAA---TACAGAACCAGGCATCG 2761
QY 111 gLeuProGluGlyHis-----ProLeuLeuGluLysArgAlaGluTyrPheArgHisLe 129
Db 2760 GCACCCCATCTCATCTGTGGCTGTGACCAAGCTGTGGCCACGCCGGGCACCT 2701
QY 129 uArgSerLeuLysSer 134
Db 2700 TCGGGCACTTGGAGT 2695

RESULT 6
AX430339/c
LOCUS AX430339 4839 bp DNA linear PAT 28-JUN-2002
DEFINITION Sequence 121 from Patent WO0240715.
ACCESSION AX430339
VERSION AX430339.1 GI:21655703
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.
TITLE Molecules for disease detection and treatment
JOURNAL Patent: WO 0240715-A 121 23-MAY-2002;
INCYTE GENOMICS INC (US)
FEATURES
Location/Qualifiers
source
1..4839
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/notes="Incype ID No: LG-233258.3:2000SEP08"

BASE COUNT 946 a 1467 c 1418 g 1008 t
ORIGIN
Alignment Scores:
Pred. No.: 225 Length: 4839
Score: 92.50 Matches: 38
Percent Similarity: 46.58% Conservative: 30
Best Local Similarity: 26.03% Mismatches: 52
Query Match: 13.06% Indels: 26
DB: 6 Gaps: 6

US-10-087-573-2 (1-141) x AX430339 (1-4839)
QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22
Db 2839 GCAGCACCACCACTCCCAAGGAGGTGCCACGGCCCTACTAATAGC----- 2789
QY 23 AspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42
Db 2788 -----AGGTTGAGTTCAGTGCCAGGCCAGCGGTGG 2753
QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeu-----ProGlyThrGlyAlaSerAla 60
Db 2752 GTGATATCCAGGTACAGCACTGTGGCACTCCGGACACGAGGAGGCCCCGTGCT 2693
QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgProGln 80
Db 2692 ATCACCTCCAGCTCAGCTCCGCTGCTTCCGCGGGTCCGCGCCCAACCCACAGAG 2633
QY 81 ---SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArg----- 95
Db 2632 GTGGCTGTTCCGCTGCTGCTGCCCACTGTCCACCGCAGGTACAGGCTCTGTAGTC 2573
QY 96 -----GluMetAsnLysThrIleSerGlnGluSerAlaArgVal-AsnHisAr 111
```

```
Db 2572 TGGTTAATACCAAAATAGGGGAAGAGTGGCAAGGGAA---TACAGAACCAGGCCATCG 2516
QY 111 gLeuProGluGlyHis-----ProLeuLeuGluLysArgAlaGluTyrPheArgHisLe 129
Db 2515 GCACCCCATCTCATCTGTGGCTGTGACCAAGCTGTGGCCACGCCGGGCACCT 2456
QY 129 uArgSerLeuLysSer 134
Db 2455 TCGGGCACTTGGAGT 2440

RESULT 7
AB024056/c
LOCUS AB024056 5973 bp DNA linear BCT 27-FEB-1999
DEFINITION Streptomyces fradiae gene for beta-lactamase and regulator protein, complete cds.
ACCESSION AB024056
VERSION AB024056.1 GI:4514646
KEYWORDS DNA-binding protein; regulator protein; beta-lactamase.
SOURCE Streptomyces fradiae
ORGANISM Streptomyces fradiae
REFERENCE 1
AUTHORS Ogawara,H. and Urabe,H.
TITLE Streptomyces fradiae beta-lactamase regulator proteins Published Only in Database (1999)
JOURNAL
REFERENCE 2
AUTHORS Ogawara,H. and Urabe,H.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-1999) Hiroshi Ogawara, Meiji Pharmaceutical University, Department of Biochemistry; Noshio, 2-522-1, Kiyose, Tokyo 204-8588, Japan (E-mail:hogawara@my-pharm.ac.jp, Tel:81-424-95-8474, Fax:81-424-95-8474)
FEATURES
Location/Qualifiers
source
1..5973
/organism="Streptomyces fradiae"
/mol_type="genomic DNA"
/strain="Y59"
/db_xref="taxon:1906"
1737..2156
/notes="putative"
/codon_start=1
/transl_table=11
/product="regulator protein"
/protein_id="BAA75484.1"
/db_xref="GI:4514647"
/translation="MATGPPAHGADDLTFRFTGALEDVTAARLVTDRYLDALERAAR PSEPRDRDVLVVELAGNAVQVPGPFSRLRLRTFDGVHVVVRDNPVPVPRPWS PADNTGAVGWLYVNALATQNVNTRPEGKDVHFLPW"
complement (1838..2128)
/notes="putative"
/codon_start=1
/transl_table=11
/product="DNA-binding protein"
/protein_id="BAA75485.1"
/db_xref="GI:4514648"
/translation="MPSGRVSTFTCVASAFTRYQPTAPVLSAGDQGRGTGGRGLSRT TTWTPSNVRSRRRNGPCAYWTALPASSVTSSSRSSSLGRAARSAR"
2539..3279
/notes="putative"
/codon_start=1
/transl_table=11
/product="regulator protein"
/protein_id="BAA75486.1"
/db_xref="GI:4514649"
/translation="MRRLGARGHGLGADGPPPGADVPRPGAPSSLTLPYGCAGAG SGGDGVTAETKTPARRAVSNTVEPRARSRLPVRTERASGVAVDTAPARALIS RSRASGPTGSGSRRAACACRRLVAGTVAASAAASSTVPTPSQSSVRRYSG GSSRARCTGAGGPGRRGRNRHROGASACAKARDTQPTGSGSWKTSQCCSHARK ACSTESKASPRRPVSRVYTWIRAGAARS"
complement (2568..3488)
/notes="unknown protein"
CDS
```

```

/codon_start=1
/transl_table=1
/protein_id="BAA75487.1"
/db_xref="GI:4514650"
/translation="MDAETTAEPVLPDTSAAFGSGEAPATGPDAAEPGDDAAPSAG
ATAACGGGAAPVACVGRVEFDALYERAAALIHVYLLTGRGGDAPDSVEHAFQR
AWEHWPVDEGLDPGVWRSEAEHAPALPWRPERRPGRPPADPVHRAALLPLLY
RTVLLCGDGVGLTVDEAAAEAAATVPAITSLERHARALLERLPFVGPPEARELSA
LAGAVSTATLPLAKSVTSGERRLALTGVFLGTALLAGAVVFSAVTPRSPSPFAP
APYGSVDDGAPGGTAPPGGSPAPPSP"
4317.5237
/notes="see also D13898"
/codon_start=1
/transl_table=1
/product="beta-lactamase"
/protein_id="BAA75488.1"
/db_xref="GI:4514651"
/translation="MDRTTAPRNRRAVLATGVGAALATAAAGCPAHAPGGRGVRGE
RLALERTHADLGAFAFDGTGTAVYADERFPPIASMEFTIYAAVLRDLDRDGEV
LARVHYTADYVKSQSPVPTGVNVAAGMTVAELCEATLTSNDTAAALLLHDLGG
PVTATRCRSVGDHVTLDLWPELNSAEPGRVTDITSPRAIGRTYGRLLILGLLAAH
DRERLTWMLDNRKTSDBRFRKGLPADLLADKTTGGDYGTTNDAGVAMPGRFPVLA
VQTRFTPDAAEDNVLAARLLAEMTD"

```

CDS

```

BASE COUNT      693 a  2192 c  2450 g   638 t
ORIGIN

Alignment Scores:
Pred. No.:      288      Length:      5973
Score:          92.50      Matches:      39
Percent Similarity: 39.44%      Conservative: 17
Best Local Similarity: 27.46%      Mismatches: 55
Query Match:      13.06%      Indels:      31
DB:              1          Gaps:      4

US-10-087-573-2 (1-141) x AB024056 (1-5973)

```

```

QY 7 ThrThrAsnPhaValAlaGluAsnArgProThrPheGlyGlu-ThrPheAspValMetAr 26
Db 1468 ACCGTGACGTGGTGGCTCTCGGGAGCGTACGGCAGGAGGACCGATCGGTACG 1409
QY 26 g-GluAlaLeuArgValLys-----SerSerGluArgLeuA 39
Db 1408 ACGAGTTCCTCACGGGGTGGCTGAGCGGGCGGAGTACTACGACGAGGAGCGCGC 1349
QY 39 laMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlas 59
Db 1348 ACATCACCGCGTGGGTGCTGGGGGTGCTCGGCCACCGCGTGGCGGAGCGCGC 1289
QY 59 erAlaIleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgP 79
Db 1288 AC-----CTGGCCCGGTGCTGCGGAGCGGCGTCTCGCGAGCTCTCGCCCGCGGTG 1235
QY 79 roGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle-----94
Db 1234 ACCGCGCGCGCGGTGCTGGTATCGAGGAGTTCACGCGCGCGTCTCGCGATGCGCACCG 1175
QY 95 -----ArgGluMetAsnLysThrIle-----101
Db 1174 GACCGGTCTCTGATACCGCGGAGTGGGACCGGACCGCGTCTGACACCGCTGGCGAGA 1115
QY 102 -----SerGlnGluSerAlaArgValAsnHisArgLeuProGlyGlyHisProLeuL 119
Db 1114 TGGTCCCGCGCGAGGAGTGGACCGGATCGTGGCGCGGTCTCGCGCGGTCTCGCGCGC 1055
QY 119 eu 119
Db 1054 TC 1053

RESULT 8
AB027763/c
LOCUS
DEFINITION Streptomyces fradiae gene, partial cds.
ACCESSION AB027763

```

```

VERSION      AB027763.1  GI:4958945
KEYWORDS
SOURCE      Streptomyces fradiae
ORGANISM    Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1 (bases 1 to 6573)
AUTHORS     Ogawara,H.
TITLE       Streptomyces
JOURNAL     Published Only in DataBase (1999)
REFERENCE   2 (bases 1 to 6573)
AUTHORS     Ogawara,H.
TITLE       Direct Submission
JOURNAL     Submitted (26-MAY-1999) Hiroshi Ogawara, Meiji Pharmaceutical
            University, Department of Biochemistry; Noshio,2-522-1, Kiyose,
            Tokyo 204-8588, Japan (E-mail:hogawara@my-pharm.ac.jp,
            Tel:81-424-95-8474, Fax:81-424-95-8474)
FEATURES
            source      1..6573
                        /organism="Streptomyces fradiae"
                        /mol_type="genomic DNA"
                        /strain="Y59"
                        /db_xref="taxon:1906"
                        complement(<1..1310)
                        /note="product is unknown
                        the 600th nucleotide of this sequence is corresponding to
                        the 1st nucleotide of the sequence of acc#AB024056"
                        /codon_start=1
                        /transl_table=11
                        /protein_id="BAA78104.1"
                        /db_xref="GI:4958946"
                        /translation="MARPADSAGVTGTPFAGTDAQGRVGRFVDAHHNIMSNEAFGG
                        RLICGKPSERGVADALDCEPHYDGLALPDFITRGDGRHDPDGWTFFADWPAHD
                        SLTHQONYVATERALARGQVRLVNDLVNTGVCVYFFKDRGCCDEMSIRLOAARTY
                        ENQEVDRMYGGGKGWFRIVTDSQAQREVVRQRLAVLVGVETSEPGCKQYLVAR
                        CEEDIDGLDLRLHLYRSMELCHKFNALCGVDFSGALGTAVNVGGLSTGTFWR
                        TPCPTQPDNDPFGGASPAKLPVGVTPAYAGARCNTRGJTLDLGAYVRGNMR
                        ROMLEIDHMSYAAAGRALDILESEAYVLSHSHSMDAEMTERVYRLGGFIAQYMHG
                        SQAQTAERARTDALRYGVGYGTDMNGVGGWPAPRGADAPD"
BASE COUNT      770 a  2396 c  2685 g   722 t
ORIGIN

Alignment Scores:
Pred. No.:      323      Length:      6573
Score:          92.50      Matches:      39
Percent Similarity: 39.44%      Conservative: 17
Best Local Similarity: 27.46%      Mismatches: 55
Query Match:      13.06%      Indels:      31
DB:              1          Gaps:      4

US-10-087-573-2 (1-141) x AB027763 (1-6573)

```

```

QY 7 ThrThrAsnPhaValAlaGluAsnArgProThrPheGlyGlu-ThrPheAspValMetAr 26
Db 2068 ACCGTGACGTGGTGGCTCTCGGGAGCGTACGGCAGGAGGACCGATCGGTACG 2009
QY 26 g-GluAlaLeuArgValLys-----SerSerGluArgLeuA 39
Db 2008 ACGAGTTCCTCACGGGGTGGCTGAGCGGGCGGAGTACTACGACGAGGAGCGCGC 1949
QY 39 laMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlas 59
Db 1948 ACATCACCGTGGGTGCTGGGGGTGCTCGGCCACCGCGTCTCGCGCGCGGCGCGC 1889
QY 59 erAlaIleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgP 79
Db 1888 AC-----CTGGCCCGGTGCTGCGGAGCGGCGTCTCGCGAGCTCTCGCCCGCGGTG 1835
QY 79 roGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle-----94
Db 1834 ACCGCGCGCGCGGTGCTGGTATCGAGGAGTTCACGCGCGCGGTCTGGGATGCCACCG 1775
QY 95 -----ArgGluMetAsnLysThrIle-----101

```

```

Db      1774 GACCGGCTCCTGTTACCGCGAGTGGACGACGCCCGCTCTGACACGCTGGCGCGAGA 1715
Qy      102 -----SerGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuL 119
Db      1714 TGGTGGCGCGGAGGAGTGGACGCGATCGTGGCGGCGCTCCCGACGCGCTACGCGCGGC 1655
Qy      119 eu 119
Db      1654 TC 1653

RESULT 9
AX367101/c      10123 bp      DNA      linear      PAT 16-FEB-2002
LOCUS
DEFINITION Novel human datus gene and protein encoded thereby.
ACCESSION AX367101
VERSION AX367101.1 GI:18955303
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Griffin,J.A., Kallick,D.A., Tribouley,C.M., Yue,H., Nguyen,D.B.,
Tang,Y.T., Lal,P., Policky,J.L., Azimzai,Y., Lu,D.A., Graul,R.,
Yao,M.G., Burford,N., Hatalla,A.J., Baughn,M.R., Bandman,O.,
Patterson,C., Yang,J., Xu,Y., Warren,B.A., Ding,L. and
Sanjanwala,M.S.
TITLE Receptors
JOURNAL Patent: WO 0198354-A 20 27-DEC-2001;
Incyte Genomics, Inc. (US)
FEATURES
source
1..10123
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/Note="Incyte ID No: 1897612CB1"
BASE COUNT 1935 a 3207 c 2984 g 1997 t
ORIGIN
Alignment Scores:
Pred. No.: 537 Length: 10123
Score: 92.50 Matches: 38
Percent Similarity: 46.58% Conservative: 30
Best Local Similarity: 26.03% Mismatches: 52
Query Match: 13.06% Indels: 26
DB: 6 Gaps: 6

US-10-087-573-2 (1-141) x AX367101 (1-10123)
Qy      3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22
Db      8612 GCAAGCACCACCACTCCCAAGGAGGTGCCACGGCCCTACTAATAGC----- 8562
Qy      23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42
Db      8561 -----AGGTTGAGGTGAGGTGCCAGGCCACCGAGTCCGGTGTGG 8526
Qy      43 AlaLeuAlaGlyMetCysGlyHisArgValLeu-----ProGlyThrGlyAlaSerAla 60
Db      8525 GTGATATCCACGGTCACAGGCACTGGGCACCTCCGGACACGAGGAGGCCCGCTGCT 8466
Qy      61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProArgProGln 80
Db      8465 ATACCTCCAGGCTCAGCTCCCGTGTGCTTCCGCGGGTCCGCGCCCGCCACCCAGAG 8406
Qy      81 ---SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArg----- 95
Db      8405 GTGGCTGTTCCGCTGCTGGTCCCGACTGTCCACCCGAGTACAGGCTCCTGTATGC 8346
Qy      96 -----GluMetAsnLysThrIleSerGlnGluSerAlaArgVal-AsnHisAr 111
Db      8345 TGGTTAATACCAATAATAGGGGAAGAGGTGGCAGGGAA---TACAGAACCAGGCGCATCG 8289

```

```

Qy      111 gLeuProGluGlyHis-----ProLeuLeuGluLysArgAlaGluThrPheArgHisLe 129
Db      8288 GCACCCCATCTCTCATCTGTGGCTGCACGTGACCAAGAGTGTGGCCACCGCGGCACCT 8229
Qy      129 uArgSerLeuLysSer 134
Db      8228 TCGGGCACCTTGGAGT 8213

RESULT 10
BD084093/c      10759 bp      DNA      linear      PAT 27-AUG-2002
LOCUS
DEFINITION Novel human datus gene and protein encoded thereby.
ACCESSION BD084093
VERSION BD084093.1 GI:22629703
KEYWORDS JP 2001327295-A/1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Nakayama,M., Nagase,T., Ohara,O. and Nakajima,D.
JOURNAL Patent: JP 2001327295-A 1 27-NOV-2001;
KAZUSA DNA RESEARCH INSTITUTE
COMMENT OS Homo sapiens (human)
PN JP 2001327295-A/1
PD 27-NOV-2001
PF 27-JUL-2000 JP 2000226291
PI MANABU NAKAYAMA, TAKAHIRO NAGASE, OSAMU OHARA, DAISUKE NAKAJIMA
PC C12N15/09, C07K14/47//A61K31/711.A61K38/00,A61K39/395,A61K48/00, PC
A61P43/00,
PC C12N15/00,A61K37/02
CC Novel human datus gene and protein encoded thereby FH Key
Location/Qualifiers
FT CDS Location/Qualifiers
source
1..10759
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 2008 a 3412 c 3256 g 2083 t
ORIGIN
Alignment Scores:
Pred. No.: 577 Length: 10759
Score: 92.50 Matches: 38
Percent Similarity: 46.58% Conservative: 30
Best Local Similarity: 26.03% Mismatches: 52
Query Match: 13.06% Indels: 26
DB: 6 Gaps: 6

US-10-087-573-2 (1-141) x BD084093 (1-10759)
Qy      3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22
Db      9265 GCAGCACCACCACTCCCAAGGAGGTGCCACGGCCCTACTAATAGC----- 9215
Qy      23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42
Db      9214 -----AGGTTGAGGTGAGGTGCCAGGCCACCGAGTCCGGTGTGG 9179
Qy      43 AlaLeuAlaGlyMetCysGlyHisArgValLeu-----ProGlyThrGlyAlaSerAla 60
Db      9178 GTGATATCCACGGTCACAGGCACTGGGCACCTCCGGAAACGAGGAGGCCCGCTGCT 9119
Qy      61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProArgProGln 80
Db      9118 ATCACTCCAGCTCAGCTCCCGTGTGCTTCCGCGGGTCCGCGCCCGCCACCCAGAG 9059
Qy      81 ---SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArg----- 95

```

Db 9058 GTGGCTGTTCCGTGCTGTGGTGCAGCAGTCTCCACCGCAGGTACAGGGCTCTCTGTAGTC 8999

QY 96 -----GluMetAenLysThrIleSerGlnJusSerAlaArgVal-AsnHisAr 111  
          :::     :::     :::     |||     |||     |||     |||     |||  
Db 8998 TGGTTAATCAAAATAGGGGAAGAGGTGGCAAGGAA---TACAGAACCAGGCCATCG 8942

QY 111 gLeuProGlUGlyHis-----ProLeuLeuGLyLysArgAlaGluTyR-PheArgHisLe 129  
      |||     |||     |||     |||     |||     |||     |||     |||  
Db 8941 GCACCCCCCATCTTCATCTGTGGCTGCACGTGACCCAAGCTGTGGCACCGCCGGCACCT 8982

QY 129 uArgSerLeuLysSer 134  
      |||     |||     |||     |||     |||     |||     |||     |||  
Db 8881 TCGGGCACITGGAGT 8866

RESULT 11  
AB053446/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

AB053446 10759 bp mRNA linear PRI 15-NOV-2001  
Homo sapiens mRNA for KIAA1773 protein (dachsous homologue),  
complete cds.  
AB053446 GI:14625440  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 Nakajima,D., Nakayama,M., Kikuno,R., Hirose,M., Nagase,T. and Ohara,O.  
Identification of three novel non-classical cadherin genes through comprehensive analysis of large cDNAs  
Brain Res. Mol. Brain Res. 94 (1-2), 85-95 (2001)

2 (bases 1 to 10759)  
Nakajima,D., Nakayama,M. and Ohara,O.  
Direct Submission  
Submitted (10-JAN-2001) Daisuke Nakajima, Kazusa DNA Research Institute, Laboratory for Genome Informatics; Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:nakajima@kazusa.or.jp, Tel:81438523915, Fax:81438523914)

FEATURES  
source  
location/Qualifiers  
1..10759  
    organism="Homo sapiens"  
    /mol\_type="mRNA"  
    /db\_xref="taxon:9606"  
    /chromosome="11"  
    /map="11q15.5"  
    (clone="hj00752")  
    /tissue\_type="brain"  
    /dev\_stage="adult"  
    /note="vector:pluescript II sk plus"  
2..10759  
    gene="KIAA1773"  
    translations="MOKELCIVSPCMKSPRHLLPLLLLLLLLILLAGVPAGWQA  
GSTDLQIDBOPQTALIGDTSAGLPAGTAAPLMYFISAOEGSGVGTDLAIDEHSGVVR  
TAYVLREQDRPRFTAVTPDGATVEYTVRADINHAFPAFPORRALQPVERTAGTGT  
RYLEPARADARLGRTQGTFALSDGAGETFRLETPTPGDTDPVELVTGELDRNR  
SHYMLQLEADYGGSPPRAQLADLVTLINDIHAFNQSRHYAVVSLSAPGSPVLQ  
VFASDAAGYGVAVTYINRRSQGPFSIDAHTGLQLERPLDFEQRRVHELVVQA  
RDGAPHGLSVGAVTHVDANDNQSMVIFLSADGPSQVFAAPPGLVARISVD  
PDGGDFAHVNVSLGEQGHFALSTQDSVIYVCVARLDREEDRNLRVATDSSGP  
PLAEAFVLHVTDVNDAFAFDRLYREPELVALPGSFVYRVARTARDPQGTNGQV  
TYSLAGAHNTWSIDPTSGIIITYASLDYLEPEPQLIVATDGGFLPASATVSV  
ALQOVNDNEPQOFRTYNALSPGTQTCFLQVATDADSGPFGLLSYSLGALGVS  
GSPFFRI DAHSQGVTRTLDRDQGSSEDFVTVAVDGGLSKMWYKVFLSDENPN  
POPYPREYASIQACQPPGTVALLRAHDPDQSGHRLGYHLILAGNSPPLFTLDEQS



QY 81 ---SerThylsSerProGluLeuArgGluLeuSerArgYsilleArg----- 95  
Db 9058 GTGGCTGTTCCGCTGCTGTCACGACGTGTCACCCGAGTACAGGCTCCTCTAGTC 8999  
QY 96 -----GluMetAsnYsThrIleSerGlnGlnSerAlaArgVal-AsnHisAr 111  
Db 8998 TGGTTAATACCAAAATAGGGGAAGGTTGGCAAGGAA---TACAGAACCAAGGCATCG 8942  
QY 111 gLeuProGluGlyHis-----ProLeuLeuGluGlyArgAlaGluTyPheArgHisLe 129  
Db 8941 GCACCCCACTCTCTCTGTCGCTGCACGTGACCAAGCTGTGCGCACCGCGGCACCT 8882  
QY 129 uArgSerLeuLysSer 134  
Db 8881 TCGGGCACTTGGAGT 8866

RESULT 12  
AC009796 192239 bp DNA linear PRI 27-MAR-2001  
LOCUS Homo sapiens chromosome 11, clone RP11-211E17, complete sequence.  
DEFINITION AC009796  
ACCESSION AC009796  
VERSION AC009796.6 GI:13324838  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 192239)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens chromosome 11, clone RP11-211E17  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 192239)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,  
Castle,A., Cerny,J., Collangelo,M., Collins,S., Collymore,A.,  
Cooke,P., DeArelano,K., Depayre,E., Devon,K., Dewar,K.,  
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,  
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,  
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kam,L.,  
Karacas,A., Lebecky,J., Lieu,C., Locke,K., Macdonald,P.,  
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,  
Melidrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,  
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,  
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Savery,P.,  
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,  
Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,  
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.  
Direct Submission  
Submitted (01-SEP-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 192239)  
REFERENCE  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Boguslavskiy,L., Boukhalter,B., Brown,A.,  
Camrata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,  
Collins,S., Collymore,A., Cooke,P., DeArelano,K., Dewar,K.,  
Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,  
Iliev,I., Johnson,R., Jones,C., Karatas,A., Lakocque,K.,  
Lamazares,R., Landers,T., Lebecky,J., Levine,R., Liu,G.,  
McLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,  
McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,  
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,  
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,  
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,  
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,  
Severy,P., Sougnuez,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Tesfaye,S.,  
Theodore,J., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission  
JOURNAL Submitted (27-MAR-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Mar 14, 2001 this sequence version replaced gi:12958053.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L2162  
Center clone name: 211\_E\_17  
-----

FEATURES  
source  
1..192239  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/map="11"  
/clone="RP11-211E17"  
/clone\_lib="RPC1-11 Human Male BAC"  
2608..2856  
/rpt\_family="MIR"  
complement(2843..3155)  
/rpt\_family="L2"  
complement(3383..3472)  
/rpt\_family="MIR"  
complement(3457..4176)  
/rpt\_family="L2"  
4588..4665  
/rpt\_family="L2"  
complement(4670..4721)  
/rpt\_family="MIR"  
complement(4741..5054)  
/rpt\_family="AluSx"  
5081..5535  
/rpt\_family="L2"  
5596..6147  
/rpt\_family="L2"  
6166..6246  
/rpt\_family="MIR"  
complement(6247..6562)  
/rpt\_family="AluJb"  
6563..6626  
/rpt\_family="MIR"  
6628..6758  
/rpt\_family="FLAM\_A"  
6776..7085  
/rpt\_family="AluSg"  
7088..7135  
/rpt\_family="AT-rich"  
7136..7247  
/rpt\_family="MIR"  
complement(7544..7681)  
/rpt\_family="L2"  
complement(7794..7875)  
/rpt\_family="L2"  
7977..8441  
/rpt\_family="L2"  
8442..8546  
/rpt\_family="L1MA4A"  
8547..8561  
/rpt\_family="L2"  
8562..8862  
/rpt\_family="AluSx"  
8863..8970  
/rpt\_family="L2"  
9753..9887  
/rpt\_family="MIR"

Alignment Scores:  
Pred. No.: 1.72e+04 Length: 192239  
Score: 92.50 Matches: 38  
Percent Similarity: 46.58% Conservative: 30  
Best Local Similarity: 26.03% Mismatches: 52  
Query Match: 13.06% Indels: 26  
DB: 9 Gaps: 6

US-10-087-573-2 (1-141) x AC009796 (1-192239)

QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
Db 166656 GCAAGCACCACCAACACTCCCAAGAGGCTGCCACGCCCTACTAATAGC----- 166706  
QY 23 AspValMetArgGluAlaLeuLeuArgVallySerSerGluArgGluAlaMetLeuArg 42  
Db 166707 -----AGGTGAGGTACAGGTGCCAGCCAGTCGGTGTGG 166742  
QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeu-----ProGlyThrGlyAlaSerAla 60  
Db 166743 GTGATATCCACGGTCACAGGCACCTGTCGCACTCCGGGAACCCAGGCAGAGCCCGGTGCT 166802  
QY 61 IleAlaAlaThrValThrProGlyGlyAlaSerMetIleuLeuLysProArgProGln 80  
Db 166803 ATCACTCCAGCTCAGCTCCCGTGGTCTTCCCGCGGGTCCGGCCACCCACCCAGAG 166862  
QY 81 ---SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArg----- 95  
Db 166863 GTGGCTGTTCGCTGCCTGTCGTCGCCGACTGCCACCCAGGTACAGGCTCCTGTAGTC 166922  
QY 96 -----GluMetAsnLysThrIleSerGlnGluSerAlaArgVal-AsnHisAr 111  
Db 166923 TGGTTAATACCAATAGGGGAGAGGTGCCAAGGAA---TACGAACACAGGCCATCG 166979  
QY 111 GluProGluGlyHis-----ProLeuLeuGluLysArgAlaGluTyrPheArgHisLe 129  
Db 166980 GCACCCCATCTCATCTGTGGCTGTCACGTGACCCCAAGCTGTGCCACGCCGGCACCT 167039  
QY 129 uArgSerLeuLysSer 134  
LJ 167040 TCGGGCACTTGAAGT 167055

AC027641 208430 bp DNA linear HTG 07-JUL-2000  
Homo sapiens chromosome 11 clone RP11-732A19, WORKING DRAFT  
SEQUENCE, 18 unordered pieces.  
AC027641 AC027641.2 GI:8570385  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Waterston,R.H. 1 (bases 1 to 208430)  
The sequence of Homo sapiens clone  
Unpublished 2 (bases 1 to 208430)  
Waterston,R.H.  
Direct Submission  
TITLE University (30-MAR-2000) Genome Sequencing Center, Washington  
SUBMITTED (30-MAR-2000) School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Jun 17, 2000 this sequence version replaced gi:7344780.  
----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
----- Project Information -----  
Center project name: y NH0232A10

RESULT 13  
AC027641/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
JOURNAL  
TITLE  
REFERENCE  
AUTHORS  
JOURNAL  
COMMENT

```
----- Summary Statistics -----
Sequencing vector: M13; 98%
Chemistry: Dye-terminator ET; 98% of reads
Chemistry: Dye-terminator Big Dye; 2% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 198825 bases at least Q40
Consensus quality: 202103 bases at least Q30
Consensus quality: 203934 bases at least Q20
Insert size: 219000; agarose-fp
Insert size: 206730; sum-of-contigs
Quality coverage: 4.27 in Q20 bases; agarose-fp
Quality coverage: 4.55 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1760: contig of 1760 bp in length
1761 1860: gap of unknown length
1861 3171: contig of 1311 bp in length
3171 3271: gap of unknown length
3271 5628: contig of 2357 bp in length
5628 5729: gap of unknown length
5729 7760: contig of 2032 bp in length
7760 7861: gap of unknown length
7861 15503: contig of 7643 bp in length
15503 15604: gap of unknown length
15604 18871: contig of 3268 bp in length
18871 22431: gap of unknown length
22431 22530: contig of 3459 bp in length
22530 28556: contig of 6026 bp in length
28556 28557: gap of unknown length
28557 35414: contig of 6758 bp in length
35414 35515: gap of unknown length
35515 42738: contig of 7224 bp in length
42738 42839: gap of unknown length
42839 50407: contig of 7469 bp in length
50407 62974: contig of 12567 bp in length
62974 63074: gap of unknown length
63074 73753: contig of 10679 bp in length
73753 73854: gap of unknown length
73854 87631: contig of 13777 bp in length
87631 106803: contig of 19072 bp in length
106803 106902: gap of unknown length
106902 120751: contig of 13748 bp in length
120751 156637: contig of 35887 bp in length
156637 156738: gap of unknown length
156738 208430: contig of 51693 bp in length.
-----
Location/Qualifiers
1..208430
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="11"
/clone="RP11-732A19"
-----
misc_feature 1..1760
/notes="assembly_name:Contig8"
misc_feature 1861..3171
/notes="assembly_name:Contig9"
misc_feature 3272..5628
/notes="assembly_name:Contig10"
misc_feature 5729..7760
/notes="assembly_name:Contig11"
misc_feature 7861..15503
```

```
/note="assembly_name:Contig12"
15604..18871
/notes="assembly_name:Contig13
clone end:T7
vector_side:right"
misc_feature 18972..22430
/notes="assembly_name:Contig14"
22531..28556
/notes="assembly_name:Contig15"
28657..35414
/notes="assembly_name:Contig16"
35515..42738
/notes="assembly_name:Contig17"
42839..50307
/notes="assembly_name:Contig18"
50408..62974
/notes="assembly_name:Contig19"
63075..73753
/notes="assembly_name:Contig20"
73854..87630
/notes="assembly_name:Contig21"
87731..106802
/notes="assembly_name:Contig22"
106903..120650
/notes="assembly_name:Contig23
clone end:SP6
vector_side:left"
misc_feature 120751..156637
/notes="assembly_name:Contig24"
156738..208430
/notes="assembly_name:Contig25"
-----
BASE COUNT 59749 a 44894 c 43435 g 58647 t 1705 others
ORIGIN
-----
Alignment Scores:
Pred. No.: 1.89e+04 Length: 208430
Score: 92.50 Matches: 38
Percent Similarity: 46.58% Conservative: 30
Best Local Similarity: 26.03% Mismatches: 52
Query Match: 13.06% Indels: 26
DB: 2 Gaps: 6
-----
US-10-087-573-2 (1-141) x AC027641 (1-208430)
QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22
Db 153648 GCAAGCACCAACCACTCCACAGAGGCTGCCACGCCCTACTAATAGC----- 153598
QY 23 AspValMetArgGluAlaLeuLeuArgVallySerSerGluArgLeuAlaMetLeuArg 42
Db 153597 -----AGGTTGAGGTCAGGTGCCAGGCCAGTCCCGGTGGTGG 153562
QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeu-----ProGlyThrGlyAlaSerAla 60
Db 153561 GTGATATCCACGGTCACAGGCACCTGTCGCACTCCGGGAACCCAGCAGAGGCCCGGTGCT 153502
QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgProGln 80
Db 153501 ATCACCCTCCAGCTCAGCTCCCGGTGGTCTCCCGCGGGTCCGGCCCGCCAGAG 153442
QY 81 ---SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArg----- 95
Db 153441 GTGGCTGTTCGGCTGCTGTGTGCCCGACTGTCCACCCCGCAGGTACAGGCTCCTGTAGTC 153382
QY 96 -----GluMetAsnLysThrIleSerGlnGluSerAlaArgVal-AsnHisAr 111
Db 153381 TGGTTAATACCAAAATAGGGGAAGAGGTGGCAAGGAA---TACAGAACCCAGCCATCG 153325
QY 111 gLeuProGluGlyHis-----ProLeuLeuGluLysArgAlaGluThrPheArgHisLe 129
Db 153324 GCACCCCATCTCTATCTGTGGCTGCACGTGACCCCAAGCTGTGGCCACGCCCGGCACCT 153265
QY 129 uArgSerLeuLysSer 134
```

DB 153264 TCGGCACTTGAAGT 153249  
||||:|||||

## RESULT 14

AC091564

## LOCUS

AC091564

## DEFINITION

Homo sapiens chromosome 11, clone RP11-732A19, complete sequence.

## ACCESSION

AC091564.12

## VERSION

GI:22657505

## KEYWORDS

HTG.

## SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens

## REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1. (bases 1 to 211735)

Birren,B., Nussbaum,C. and Lander,E.

Homo sapiens chromosome 11, clone RP11-732A19

Unpublished

2. (bases 1 to 211735)

Birren,B., Linton,L., Nussbaum,C., Lander,E., Allen,N., Anderson,S.,

Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A.,

Canarata,J., Campopiano,A., Chang,J., Choepey,Y., Colangelo,M.,

Collins,S., Collymore,A., Cooke,P., DeArelano,K., Dewar,K.,

Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,

Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRoque,K.,

Lanazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,

MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,

McGowan,P., McKernan,K., McPheters,R., Meldrum,J., Meneus,L.,

Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,

Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,

Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,

Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,

Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,

Severy,P., Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

Strausman,A., Subramanian,A., Talamas,J., Tesfaye,S.,

Theodore,J., Travers,M., Travis,N., Triglio,J., Vassiliev,H.,

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,

Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (07-MAY-2001) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3. (bases 1 to 211735)

Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,

Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,

Canarata,J., Chang,J., Chazaro,B., Choepey,Y., Collymore,A.,

Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,

Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,

Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,

Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,

Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,

Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,

McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mienga,V.,

Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,

O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,

Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,

Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,

Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,

Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,

Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (13-AUG-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

4. (bases 1 to 211735)

Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,

Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,

Canarata,J., Chang,J., Chazaro,B., Choepey,Y., Collymore,A.,

Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,

Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,  
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mienga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,  
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,  
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (03-SEP-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 3, 2002 this sequence version replaced gi:22123080.

All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RW/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information  
Center project name: L11860  
Center clone name: 732\_A\_19

----- Location/Qualifiers  
1. 211735

/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/map="11"

/clone="RP11-732A19"  
/clone\_lib="RPC1-11 Human Male BAC"  
complement(61..177)  
/rpt\_family="L2"  
431..777

/rpt\_family="L2"  
1238..1278  
/rpt\_family="GC rich"  
complement(1971..3002)  
/rpt\_family="L2"  
complement(3014..3207)  
/rpt\_family="L1PA13"  
3391..3691  
/rpt\_family="AluSg"  
complement(5339..5470)  
/rpt\_family="L2"  
8518..8617  
/rpt\_family="MIR"  
9609..9671  
/rpt\_family="GC rich"  
complement(10276..10420)  
/rpt\_family="L2"  
complement(10477..10658)  
/rpt\_family="MIR"  
complement(11053..11154)  
/rpt\_family="L2"  
14007..14026  
/rpt\_family="(T)n"  
complement(16033..16140)  
/rpt\_family="MIR"  
17172..17281  
/rpt\_family="MIR"  
17735..18212  
/rpt\_family="L2"  
25342..25875  
/rpt\_family="Charliel"  
26542..26683  
/rpt\_family="MIR3"  
31972..32491

## FEATURES

source

## repeat\_region

repeat\_region

## repeat\_region

repeat\_region

## repeat\_region

repeat\_region

## repeat\_region

repeat\_region

## repeat\_region

repeat\_region

## repeat\_region

repeat\_region

## repeat\_region

repeat\_region

## repeat\_region

repeat\_region

## repeat\_region

repeat\_region

## repeat\_region

repeat\_region

## repeat\_region

repeat\_region

## repeat\_region

repeat\_region

## repeat\_region

repeat\_region

## repeat\_region

repeat\_region

## repeat\_region

repeat\_region

## repeat\_region

repeat\_region

## repeat\_region

repeat\_region

## repeat\_region

repeat\_region

```
repeat_region      /rpt_family="MER82"
32719..32809
/rpt_family="L2"
repeat_region      33689..33704
/rpt_family="L2"
repeat_region      complement(33705..34007)
/rpt_family="AluJo"
repeat_region      complement(34008..34133)
/rpt_family="FLAM_A"
repeat_region      34134..34707
/rpt_family="L2"
repeat_region      complement(34708..34920)
/rpt_family="MER20"
repeat_region      34921..35407
/rpt_family="L2"
repeat_region      35754..35896
/rpt_family="L2"
repeat_region      36102..36277
/rpt_family="L2"
repeat_region      36402..36624
/rpt_family="AluY"
repeat_region      39013..39048
/rpt_family="(CAG)n"
repeat_region      complement(39748..40578)
/rpt_family="L2"
repeat_region      40806..41691
/rpt_family="L2"
repeat_region      42071..42202
/rpt_family="L2"
repeat_region      42520..42637
/rpt_family="MIR"
repeat_region      43511..43664
/rpt_family="(CA)n"
repeat_region      complement(45047..45229)
/rpt_family="MIR"
unsure             complement(45590..45607)
repeat_region      /notes"<30 qual SNGL region"
complement(45602..46031)
/rpt_family="L2"
unsure             complement(45625..45629)
/notes"<30 qual SNGL region"
complement(45638..45642)
/notes"<30 qual SNGL region"
complement(45663..45669)
/notes"<30 qual SNGL region"
complement(45708..45712)
/notes"<30 qual SNGL region"
repeat_region      46134..46345
/rpt_family="MIR"
repeat_region      46361..46596
/rpt_family="MIR"
repeat_region      complement(46642..46678)
/rpt_family="MIR"
repeat_region      complement(47066..47185)
/rpt_family="L2"
repeat_region      47609..47644
/rpt_family="(TTCT)n"
repeat_region      complement(48727..50294)
/rpt_family="L2"

Alignment Scores:
Pred. No.:         1 93e-04
Score:             92.50
Percent Similarity: 46.58%
Best Local Similarity: 26.03%
Query Match:       13.06%
DB:                9

US-10-087-573-2 (1-141) x AC091564 (1-211735)

QY      3 SerThrSerThrThrThrThrThrValAlaGluAsnArgProThrPheGlyGluThrPhe 22
      :.....|
DU      20319 GGAAGCACCACCACTCCCAAGGAGGTCGCCAGCGCCCTACTAATAGC----- 20369

/rpt_family="MER82"
32719..32809
/rpt_family="L2"
33689..33704
/rpt_family="L2"
complement(33705..34007)
/rpt_family="AluJo"
complement(34008..34133)
/rpt_family="FLAM_A"
34134..34707
/rpt_family="L2"
complement(34708..34920)
/rpt_family="MER20"
34921..35407
/rpt_family="L2"
35754..35896
/rpt_family="L2"
36102..36277
/rpt_family="L2"
36402..36624
/rpt_family="AluY"
39013..39048
/rpt_family="(CAG)n"
complement(39748..40578)
/rpt_family="L2"
40806..41691
/rpt_family="L2"
42071..42202
/rpt_family="L2"
42520..42637
/rpt_family="MIR"
43511..43664
/rpt_family="(CA)n"
complement(45047..45229)
/rpt_family="MIR"
complement(45590..45607)
/notes"<30 qual SNGL region"
complement(45602..46031)
/rpt_family="L2"
complement(45625..45629)
/notes"<30 qual SNGL region"
complement(45638..45642)
/notes"<30 qual SNGL region"
complement(45663..45669)
/notes"<30 qual SNGL region"
complement(45708..45712)
/notes"<30 qual SNGL region"
46134..46345
/rpt_family="MIR"
46361..46596
/rpt_family="MIR"
complement(46642..46678)
/rpt_family="MIR"
complement(47066..47185)
/rpt_family="L2"
47609..47644
/rpt_family="(TTCT)n"
complement(48727..50294)
/rpt_family="L2"

Alignment Scores:
Pred. No.:         2 28
Score:             92.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match:       12.99%
DB:                6

US-10-087-573-2 (1-141) x AX530473 (1-90)

QY      1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPhe 18
      :.....|
DU      35 ATGGAGTCGACATCAACACGACCACTTTGTTGCCGAGAACCGTCCACCTTT 88

RESULT 16
AE004618
LOCUS
DEFINITION
Pseudomonas aeruginosa PAO1, section 179 of 529 of the complete
Genome.
ACCESSION
AE004618 AE004091
```

gene  
CDS

```

GHLSPVAFERHLLTIGGEWRKEELEDERSVNTAGDSARHKALFLODEIAFDPMWSLTL
GRFKHEAFGHESSPRLLYLLHLLSDVLTLAGRGVGRKAPSKQLSEYAAVGGGGR
FTIYNPDLPKPTNTSYELGADYQDGMGLKGTVFENDRLDIOTVCVARCGGGEI
RVNENVDRIIRGLELGGVDLPADRLWELNVTYLDARNRTAGORLGRSHRLANSOL
RASPVRFSAQIRTEVVGSOVAYSSNGVALPAYSLWHLSELSQKLSENLTLRGGIENL
GQGLADDDPTHTYAEPRGTFLGLVASF"
6136..9981
/locus_tag="PA1923"
6136..9981
/locus_tag="PA1923"
/notes="Protein name confidence: Class 4 (homologs of
previously reported genes of unknown function, or no
similarity to any previously reported sequences)"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG05311.1"
/db_xref="GI:9947917"
/translation="WRRSLLFCGALIASCAWAEPAQVRLSSDFVLPGKHQRLAGWA
RAGVELRLGLIGESAPGEWLDGNNLLIOTPRTPDRAQVEEALGERLQGGTOPWI
RVGGPGPGFNNIPALGGRVLGVYANGSEANRLRLEFEAVRWAGLTVDALPAPOLA
QAGFYHPADAPAFVADLYLANGSRAWSDAPRIAFILPRGALADAOQTGAIDELLERS
ERHQPLAWFDDSDPEALRKSFGADQVQALVNIQHLQNGPARRAEFLADLPVLIQT
LGRDNEADWLAAAGVAPRTAAAFGLMPETWGMSSDPLIVISALENGEFLKAWAGAE
LLKLDRLRLRLPAADKHLALMFNWHPEGEKNVAASHLNPASLRLGEALRAAGY
RVATSDSALIDTAQRLGGYRPTDLALYRDGLAASLPDYLHWFPEALPADREE
MRWMDPRRHWRALRIDGRRFVPAARLGNLLLPQPPRAGRCEAVHDSAVPDDH
LYLVAVQVREGGADALHFGHTQEWLPGKDRGLAVGYDLRLQHLIHEYQLDEGAVRSZ
DNVGEALQARRGTVASHQTPSPAPAGLYDELRLQHLIHEYQLDEGAVRSZ
QIRAVRAHNNDDLGNSAMHEDPPAFGLVHLHLKLAGSAMPLGHTPTGVAASP
EURLGTVMQOLGEFYRAJGLDPELFAADFRALREGRAYRTLRYLRDGGIAKVD
PLRQLRLARLELDQADTGLLEALLAGLRFVAPGGDPINRQVPSGRNLFAD
EADKVTAAAYAGAGAFQLLSEAEHQGRAPKLAFLWSSETMRHLGVESQAL
HALGLPRWDAGRLIALDIPVAAELGRPRVDVLQVTSVYRDQDFGMRLILAEIER
LAALDEPNGLARNQALRRRLRGEVPEAPLQRLSLRLFGNAPGDYGTGTQTLTLD
STWDDDSALAQFLQRYAYSGRDMVKLQDGNLFAQLKGVQAAILLRSSTLNG
LSTDHFEVYGLGLSLAVHLHDGASPALYIADLRQRPRTAAQFLASELRGRLYSPO
WIAHQRETAGSLMELDANLWQADRLMWRADQWQALHDTIFWMDRRLGLUAEF
FETHNPTAQAIARWAEAIRKYWDASEQTRREIARWRQAAAGAEVGAATREF
IERMAGGFGAEAAAGAAAGAGETVSGRVLEEVAPPSPGPQFWLRLAAALVLLF
AAGAVRQSVVNRPRMENRS"
9978..10454
/locus_tag="PA1924"
9978..10454
/locus_tag="PA1924"
/notes="Protein name confidence: Class 4 (homologs of
previously reported genes of unknown function, or no
similarity to any previously reported sequences)"
/codon_start=1
/transl_table=11

Alignment Scores:
Pred. No.: 659 Length: 11060
Score: 92.00 Matches: 44
Percent Similarity: 46.03% Conservative: 14
Best Local Similarity: 34.92% Mismatches: 51
Query Match: 12.99% Indels: 17
DB: 1 Gaps: 5

US-10-087-573-2 (1-141) x AE004618 (1-11060)

Qy 25 MetArgGluAlaLeuArgValLysSerSerGluArg---LeuAla----- 39
Db CTGCGGAGCAGTGTGTCGCGCGCGGAACTCGACCGCAGCTCGCGGACCGCGAA 8421
Qy 40 MetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSer 59
Db CTGCGGAGCAGTGTGTCGCGCGCGGAACTCGACCGCAGCTCGCGGACCGCGAA 8421
Qy 60 Alaile-AlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgPr 79
Db CCGATCCGCAATCCGCGAGGTACCCGCGCGGCAACC-----TGTTCGCTTCGAGGCC 8535

```

```

Qy 79 o-GlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIleArg-GluMetAsn 98
Db 8536 GACAAGGTACGACCGCGCGCGCTACGAAAGCCGCGCGCTTCGCGCAACTG--- 8592
Qy 99 LysThrIleSerGlnGluSerAlaArgValAsnHis-----ArgLeuProGluGlyHis 116
Db 8593 -----CTGGAAGCTACCGCGCGGACGAGCAGCGCGCGGAAAGCTC 8640
Qy 117 ProLeuLeuGluLysArgAlaGluTyPheArgHisLeuArgSerLeuLysSerGlnGly 136
Db 8641 GCCTTCAGCCTCTCGTCTCGGAACCATGCCACCTCGGCGCATCGTCGAAAGCCAGCGC 8700
Qy 137 ValAsnArgLeu 140
Db 8701 CTGACGCGCCCTC 8712

RESULT 17
AF416730/c
LOCUS AF416730 1386 bp mRNA linear ROD 25-OCT-2001
DEFINITION Rattus norvegicus CORO1A protein mRNA, complete cds.
ACCESSION AF416730
VERSION AF416730.1 GI:16417359
KEYWORDS SOURCE
ORGANISM Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1386)
AUTHORS Suzuki, K., Takeshita, F., Nakata, N. and Makino, M.
TITLE Molecular cloning of rat Corola
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1386)
AUTHORS Suzuki, K., Takeshita, F., Nakata, N. and Makino, M.
TITLE Direct Submission
JOURNAL Submitted (05-SEP-2001) Department of Microbiology, Leprosy
Research Center, National Institute of Infectious Diseases, 4-2-1
Aoba-cho, Higashi-murayama, Tokyo 189-0002, Japan
FEATURES
source
1..1386
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/sex="male"
/tissue_type="spleen"
1..1386
/codon_start=1
/product="CORO1A protein"
/protein_id="AA118695.1"
/db_xref="GI:16417360"
/translation="MSRQVVRSSKFRHVFQPAKADOCYEDVRVSQTTWDSGFCVNP
KFVALICEASGGGAFVLPLGKTRVDKNVPLVCGHTAPVLDTAWCPHNDNVIASGSE
DCTVMWEIPDGLVPLRFVTVLEGTTRGVIAWHPTAQNVLLSAGCDNVILWVD
EGTPRHAVFVSEKGLITGTFSRMSRQVALWTKHLEPLSLQELDTSSGVLLPFD
PDTNIVYLCGKGSSIRYFISITSEAPFLHYLSMFSSKESQSGMGYMPKRGLEVNKCEI
ARFKYLERKCEPIATVPRKSLFDQDLFPPTAGPDPAITAEWLSSGRAGPLLLSL
KGVFPVPSKRLRVRNGLDSARRKATPEPSTLSSDTSVRLEEDVRNLNIAVQLQKR
LDRLEETVQAK"
BASE COUNT 291 a 373 c 429 g 293 t
ORIGIN
Alignment Scores:
Pred. No.: 53.2 Length: 1386
Score: 91.50 Matches: 35
Percent Similarity: 42.61% Conservative: 14
Best Local Similarity: 30.43% Mismatches: 31
Query Match: 12.92% Indels: 35
DB: 10 Gaps: 6

US-10-087-573-2 (1-141) x AF416730 (1-1386)

```

QY 3 SerThrSerThrThrThrAsnPhelAlaGluAsnArgProThrPheGlyGluThrPhe 22  
 Db 383 TCCAAAGGTGACCAAGGTCGCGCAAGGCGACGACGAGGCGCCCATCCGGATCTCCAC 324  
 QY 23 AspValMetArgGluAlaLeuArgVallysserSerGluArgLeuAlaMetLeuArg 42  
 Db 323 ACCATA-----ACTGTGCAGTCTCTAGAGCCACTGCCATGACGTTG 282  
 QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
 Db 281 TCATTG-----TGTGGACACACGCGATGCTAGCAGGCGGCGAGTGTGCCCGAC 231  
 QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
 Db 230 ACCAGGGGCACATCTTGTCTACTGTCCTAGTCTTGGCTAGGGTAGCAGGAGGCC 171  
 QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysLeu 94  
 Db 170 CTCGCCCA-----LysThrIleSerGln 103  
 QY 95 ArgGluMetAsn-----LysThrIleSerGln 103  
 Db 146 AGAGCATGAATCTGGGTGACAGCGGACGAGACCTGTCCAA 102

BC002136 1607 bp mRNA linear ROD 16-APR-2003  
 MGC:7245 IMAGE:3484185, actin binding protein 1A, mRNA (cDNA clone  
 complete cds.  
 BC002136  
 MGC:7245 IMAGE:3484185, complete cds.  
 BC002136.1 GI:12805334  
 MGC.  
 Mus musculus  
 Mus musculus (house mouse)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 1607)  
 Strausberg, R.D., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
 Diatchenko, L., Marusik, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullah, S.J., Bosak, S.A., McEwan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,  
 Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E.,  
 Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 2388257  
 12477932  
 2 (bases 1 to 1607)  
 Strausberg, R.  
 Direct Submission  
 Submitted (31-JAN-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapbs-remail@nih.gov](mailto:cgapbs-remail@nih.gov)  
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
 Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www-ahgc.stanford.edu>  
 Contact: (Dickson, Mark) [mdg@paxil.stanford.edu](mailto:mdg@paxil.stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
 R. M.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAX Plate: 7 Row: 1 Column: 3.  
 Location/Qualifiers  
 1. 1607  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="MGC:7245 IMAGE:3484185"  
 /tissue\_type="Mammary tumor. WAP-TGF alpha model. 7 months  
 old, gross tissue."  
 /clone\_lib="NCI CGAP Mam5"  
 /lab\_host="DH10B"  
 /notes="Vector: pCMV-SPORT6"  
 1. 1607  
 /gene="Corola"  
 /notes="synonyms: Ciabp, p57, TACO"  
 /db\_xref="LocusID:12721"  
 /db\_xref="MGI:1345961"  
 95..1480  
 /codon\_start=1  
 /product="coronin, actin binding protein 1A"  
 /protein\_id="AAH02136.1"  
 /db\_xref="GI:12805335"  
 /db\_xref="LocusID:12721"  
 /translations="MSRQVVRSSKFRHFGQPAKADQCYEDVRVSOTTWDSGFCVNP  
 KFWALCEASGGAFILVPLGKTGRVDKXNPLVCGHTAPVLDIAWPHNDNVIASSE  
 DVTVMWVLPDGLVHPDTIYSVDMSRDGALICTSCDKRVRIEPRKGTVAEKDRPH  
 EGRPHVAVFVSSEKILTTGFSMSRQVALMDTKHEEPLSLQELDTSSGVLLPFD  
 PDNIVLCGKDGSSIRYEITSEAPFLHYLSMFSSKESQRMGYMPKRLGVNKEI  
 ARYKILHKECEPIAMTPRKSLDFQEDLYPPTAGDPALTAEWLGGRDAGPLLISL  
 KDGVPPKSRLEAVNRGLDSARRATPEPSGTSSDTVSRLIEDVRNLNAIVQKQER  
 LDRLESTVQAK"

BASE COUNT 349 a 454 c 475 g 329 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 75.2 Length: 1607  
 Score: 91.50 Matches: 35  
 Percent Similarity: 42.61% Conservative: 14  
 Best Local Similarity: 30.43% Mismatches: 31  
 Query Match: 12.92% Indels: 35  
 DB: 10 Gaps: 6

US-10-087-573-2 (1-141) x BC002136 (1-1607)  
 QY 3 SerThrSerThrThrThrAsnPhelAlaGluAsnArgProThrPheGlyGluThrPhe 22  
 Db 477 TCCAAAGGTGACCAAGGTCGCGCAAGGCGACGACGAGGCGCCCATCCGGATCTCCAC 418  
 QY 23 AspValMetArgGluAlaLeuArgVallysserSerGluArgLeuAlaMetLeuArg 42  
 Db 417 ACCATA-----ACTGTGCAGTCTCTAGAGCCACTGCCATGACGTTG 376  
 QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
 Db 375 TCATTG-----TGTGGACACACGCGATGCTAGCAGGCGGCGAGTGTGCCCGAC 325  
 QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
 Db 324 ACCAGGGGCACATCTTGTCTACTGTCCTAGTCTTGGCTAGGGTAGCAGGAGGCC 265  
 QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysLeu 94  
 Db 170 CTCGCCCA-----LysThrIleSerGln 103



Db 264 CCCCCCA-----CTGGCCTCACAGATC 241

Qy 95 ArgGluMetAsn-----LysThrIleSerGln 103

Db 240 AGAGCCATGAACCTTGGGTTGACAGCGCAGAGCCACTGTCCCAA 196

RESULT 19

AF495469/c

LOCUS Rattus norvegicus 1618 bp mRNA linear ROD 23-APR-2002

DEFINITION Rattus norvegicus tryptophane aspartate-containing coat protein

ACCESSION AF495469

VERSION AF495469.1 GI:20271116

KEYWORDS

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 1618)

AUTHORS Kohchi,C., Inagawa,H., Makino,K., Terada,H. and Soma,G.-I.

TITLE A new therapeutic strategy of mycobacterium infection by use of anti-TACO sequence

JOURNAL Unpublished

AUTHORS Kohchi,C., Inagawa,H., Makino,K., Terada,H. and Soma,G.-I.

TITLE Direct Submission

JOURNAL Submitted (26-MAR-2002) Radioisotope Center, Hiroshima University, 1-4-2 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan

FEATURES

Location/Qualifiers

1..1618

/organism="Rattus norvegicus"

/mol\_type="mRNA"

/strain="F34"

/db\_xref="taxon:10116"

/sex="female"

/tissue\_type="brain"

1..1618

/gene="Taco"

105..1490

/gene="Taco"

/codon\_start=1

/product="tryptophane aspartate-containing coat protein"

/protein\_id="AA18515.1"

/db\_xref="GI:20271117"

/translation="MSRQVRSKFRHVFQPAKQCYEDVRVSQTTWDSGFCVNP KFMALICEAGSGGAFVLPLGKTRGVKNVPLVCGHTAPVLDIAMCPHNDNVIASGE DCTVMWEIPDGLVPLPREPVVTLGHTKRVGIVAMHTAQNVLISAGCDNVILVMD VGTGAALTLGPDVHPDTIYSVDMRSRGALICTSCRDKRVVLEPKGTSGVVAEKRPH EGTSPHAFVVSQKILTTGFSMSRQVALWDTNHLSEPLSLQELDTSSGVLPPFD PDNIVLCKGSSIRYFETSEAPFLHYLSMFSKESQRMGMKRGLEVNKEI ARFKLHERKCEPIAMTVPRKSLDFQEDLYPTAGDPALTAEEWLSGRDAGPLLI SL KDGVPKPSRLRVNRLGDSARRRATPEPSSTLSSDTVSRLEEDVRLNLAIVQKLQER LDRLEETVQAK"

BASE COUNT 353 a 444 c 476 g 345 t

ORIGIN

Alignment Scores:

Pred. No.: 75.8 Length: 1618

Score: 91.50 Matches: 35

Percent Similarity: 42.61% Conservative: 14

Best Local Similarity: 30.43% Mismatches: 31

Query Match: 12.92% Indels: 35

DB: 10 Gaps: 6

US-10-087-573-2 (1-141) x AF495469 (1-1618)

Qy 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22

Db 487 TCCAAAGGTGACACAGGCTCCCGAGCGCAGCAGCCACCGGATCTCCAC 428

Qy 23 AspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42

Db 427 ACCATA-----ACTGTGAGTCTCTCAGAGCCACTGGCAATGACGTTG 386

Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58

Db 385 TCATTG-----TGGGACACACGCGATGCTTACACAGGGGCACTGTGGCCGAC 335

Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74

Db 334 ACCAGGGGCAACATCTTCTGTACTCTGCTCCAGTCTTGCCTAGGGGTAGCACAGGAAGGCC 275

Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94

Db 274 CCTCCCCA-----CTGGCCTCACAGATC 251

Qy 95 ArgGluMetAsn-----LysThrIleSerGln 103

Db 250 AGAGCCATGAACCTTGGGTTGACAGCGCAGAGCCACTGTCCCAA 206

RESULT 20

AF143955/c

LOCUS Mus musculus coronin-1 mRNA, complete cds.

DEFINITION AF143955

ACCESSION AF143955

VERSION AF143955.1 GI:4895036

KEYWORDS

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 1633)

AUTHORS Okumura,M., Kung,C., Wong,S., Rodgers,M. and Thomas,M.L.

TITLE Definition of family of coronin-related proteins conserved between humans and mice: close genetic linkage between coronin-2 and CD45-associated protein

JOURNAL DNA Cell Biol. 17 (9), 779-787 (1998)

MEDLINE 98449467

PUBMED 9778037

REFERENCE 2 (bases 1 to 1633)

AUTHORS Okumura,M., Kung,C.E., Wong,S., Rodgers,M. and Thomas,M.L.

TITLE Direct Submission

JOURNAL Submitted (19-APR-1999) Pathology, Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, MO 63110, USA

FEATURES

Location/Qualifiers

1..1633

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

138..1523

/codon\_start=1

/product="coronin-1"

/protein\_id="AAD32703.1"

/db\_xref="GI:4895037"

/translation="MSRQVRSKFRHVFQPAKQCYEDVRVSQTTWDSGFCVNP KFMALICEAGSGGAFVLPLGKTRGVKNVPLVCGHTAPVLDIAMCPHNDNVIASGE DCTVMWEIPDGLVPLPREPVVTLGHTKRVGIVAMHTAQNVLISAGCDNVILVMD VGTGAALTLGPDVHPDTIYSVDMRSRGALICTSCRDKRVVLEPKGTSGVVAEKRPH EGTSPHAFVVSQKILTTGFSMSRQVALWDTNHLSEPLSLQELDTSSGVLPPFD PDNIVLCKGSSIRYFETSEAPFLHYLSMFSKESQRMGMKRGLEVNKEI ARFKLHERKCEPIAMTVPRKSLDFQEDLYPTAGDPALTAEEWLSGRDAGPLLI SL KDGVPKPSRLRVNRLGDSARRRATPEPSSTLSSDTVSRLEEDVRLNLAIVQKLQER LDRLEETVQAK"

BASE COUNT 352 a 447 c 496 g 338 t

ORIGIN

Alignment Scores:

Pred. No.: 76.6 Length: 1633

Score: 91.50 Matches: 35

Percent Similarity: 42.61% Conservative: 14

Best Local Similarity: 30.43% Mismatches: 31

Query Match: 12.92% Indels: 35

DB: 10 Gaps: 6

US-10-087-573-2 (1-141) x AF143955 (1-1633)

```

Qy 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 520 TCAAGGTGACGACAGGCTCCGCGAGCGGACGACCGCCCATCGGATCTCCAC 461
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 23 AspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgGluAlaMetLeuArg 42
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 460 ACCATA-----ACTGTGCAGTCTCTCAGAGCCACTGGCAATGACATTG 419
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 418 TCATTG-----TGTGGACACCGCGGATGTACACAGGGCGAGTGTGGCGGAG 368
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 367 ACCAGGGGCGACGCTTCTGTACTCTGCTCCAGTCTTGCCTAGGGTAGCACAGGAAGGCC 308
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 307 CTTCCCCCA-----CTGGCTTCACAGATC 284
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 95 ArgGluMetAsn-----LysThrIleSerGln 103
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 283 AGAGCCATGAACCTTGGGTGTGACAGCGAGGAGCCACTGTCCCA 239
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 21
HSCORONIN/c HSCORONIN 1563 bp mRNA linear PRI 23-DEC-1995
LOCUS
DEFINITION H.sapiens mRNA for coronin.
ACCESSION X89109
VERSION X89109.1 GI:1136139
KEYWORDS coronin homologue.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE 1
AUTHORS Grogan,A., Keep,N.H., Reeves,E. and Segal,A.W.
JOURNAL Unpublished
PUBLISHED
2 (bases 1 to 1563)
REFERENCE 2
AUTHORS Keep,N.H.
JOURNAL Direct Submission
TITLE Submitted (23-JUN-1995) N.H. Keep, University College London,
Medicine, Rayne Institute, 5 University Street, London, WC1E 6JF,
UK
FEATURES
Location/Qualifiers
1..1563
/organism="Homo sapiens"
/mol_type="mRNA"
/strain="ATCC 8531"
/db_xref="taxon:9606"
/tissue_type="brain"
/dev_stage="infant"
/lab_host="E.coli"
94..1479
/codon_start=1
/product="coronin homologue"
/protein_id="CAA61482.1"
/db_xref="GI:1136140"
/db_xref="SWISS-PROT:P31146"
/translation="MSRQVRSKPRHVFQGPAPAKQCYEDVRVSQTTWDSGFCVAVP
KEVALICASGGAGFLVLPGLKGTGRVDKNAFTVCGHTAPVLDIAWCHNDNVIASGSE
DCTVMWEIPDGLMLPREPVLTGHTKRGIVAWHTTAQNVLSSAGCDNVLMVWD
VTGAAMTLGPEVHPDTIYSVDWSRDGLICTSCRDKRVRIIEPRKGTVAEKDRPH
EGTRPVRAVFSVSEGLITIGFSRMSERQVALWDTKLEEPSLQELDTSSGVLLPFFD
PDTNYLTCGKGDSSIRYFEITSEAPFLHYLSMFSSKESQRMGMXPRGLEVNKEI
ARFYKLERCEPIAMTVPRKSDLFQEDLYPTTAPGDPALTAEEWLGRDAGPLLISL
KDGVPVPPKSRLELVNRGLDTGRRRAAPEASGTPSSDAVSRLLEEEMRKLQATVQELQK
LDRLLETQVQAK"
300 a 492 c 494 g 277 t
BASE COUNT
ORIGIN
Alignment Scores:

```

```

Pred. No.: 98.4 Length: 1563
Score: 90.00 Matches: 40
Percent Similarity: 41.54% Conservative: 14
Best Local Similarity: 30.77% Mismatches: 41
Query Match: 12.71% Indels: 35
DB: 9 Gaps: 6

US-10-087-573-2 (1-141) x HSCORONIN (1-1563)

Qy 8 ThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg--- 26
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 494 ACAGCTGTGTGGCCCTCCAGGTGACAGCGGCTCCCGAGGGGACGATCAGGCC 435
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 27 -----GluAlaLeuLeuArgValLysSerSerGluArgGluAlaMetLeuArg 42
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 434 CCATCGGGATCTCCACACCATGACTGTGACGCTCTCGGAGCCACTGGCAATGACGTTG 375
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 374 TCATTG-----TCTGGGACACCGCGGATGTACGACAGGGGCTGTGTGGCCACAG 324
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 323 ACCGTGGGCGCATCTTGTCTCCACACGTCAGTCTTCCAGGGGACGACAGGAAGGCC 264
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 263 CTTCCCCCG-----CTGGCTTCACAGATC 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 95 ArgGluMetAsn-----LysThrIleSerGlnGlu-SerAlaArgVa 108
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 239 AGGCCCAAACTTAGGGTTGACAGCACAGAGCCACTGTCCCAGGTGGTCTGTGAGACG 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 108 LAsnHisArgLeuProGluGlyHisPro 117
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 179 CGCATCTTTCATAGCACTGTGTGGGCT 152
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 22
HSU34690/c HSU34690 1589 bp mRNA linear PRI 02-OCT-1995
LOCUS
DEFINITION Human coronin-like protein (HCO001) mRNA, complete cds.
ACCESSION U34690
VERSION U34690.1 GI:1002922
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1589)
REFERENCE 1
AUTHORS Liau,G., Popa,I., Argraves,K. and Argraves,W.S.
TITLE Direct Submission
JOURNAL Submitted (23-AUG-1995) Gene Liau, Dept. of Molecular Biology,
Holland Laboratory, American Red Cross, 15601 Crabbs Branch Way,
Rockville, MD 20855, USA
FEATURES
Location/Qualifiers
1..1589
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="kidney"
1..1589
/gene="HCO001"
/gene="HCO001"
/codon_start=1
/product="coronin-like protein"
/protein_id="AAA77058.1"
/db_xref="GI:1002922"
/translation="MSRQVRSKPRHVFQGPAPAKQCYEDVRVSQTTWDSGFCVAVP
KCFVALICEASGGAGFLVLPGLKGTGRVDKNAFTVCGHTAPVLDIAWCHNDNVIASGSE
DCTVMWEIPDGLMLPREPVLTGHTKRGIVAWHTTAQNVLSSAGCDNVLMVWD
VTGAAMTLGPEVHPDTIYSVDWSRDGLICTSCRDKRVRIIEPRKGTVAEKDRPH
EGTRPVRAVFSVSEGLITIGFSRMSERQVALWDTKLEEPSLQELDTSSGVLLPFFD
KDGVPVPPKSRLELVNRGLDTGRRRAAPEASGTPSSDAVSRLLEEEMRKLQATVQELQK
LDRLLETQVQAK"
300 a 492 c 494 g 277 t
BASE COUNT
ORIGIN
Alignment Scores:

```



REFERENCE	1	(bases 1 to 1657)
AUTHORS	Suzuki,K. and Toyoshima,S.	
TITLE	NEW PROTEIN	
JOURNAL	Patent: JP 1996119996-A 1 14-MAY-1996;	
COMMENT	JAPAN TOBACCO INC	
	OS Homo sapiens (human)	
	PN JP 1996119996-A/1	
	PD 14-MAY-1996	
	PF 21-OCT-1994 JP 1994282743	
	PI SUZUKI KENSUKE, TOYOSHIMA SATOSHI	
	PC C07K14/47, C07K16/18, C12N1/21, C12N15/09, C12P21/02, C12P21/08, PC (C12N1/21,	
	PC C12R1:19), (C12P21/02, C12R1:19);	
	CC strandedness: Double;	
	CC topology: Linear;	
	FH Key Location/Qualifiers	
	FT source 1..1657	
	FT /organism='Homo sapiens'	
	FT /cell_line='HL60'	
	FT FT CDS 100..1485	
	FT /product='a protein which has WD40 repeat and leucine zipper'.	
FEATURES	Location/Qualifiers	
source	1..1657	
	/organism="Homo sapiens"	
	/mol_type="genomic RNA"	
	/db_xref="taxon:9606"	
BASE COUNT	359 a 497 c 511 g 290 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	105	Length: 1657
Score:	90.00	Matches: 40
Percent Similarity:	41.54%	Conservative: 14
Best Local Similarity:	30.77%	Mismatches: 41
Query Match:	12.71%	Indels: 35
DB:	6	Gaps: 6
US-10-087-573-2 (1-141) x E11332 (1-1657)		
Qy	8 ThrAsnPheValAlaGluAunArgProThrPheGlyGluThrPheaspValMetArg---	26
Db	500 ACACGGTTGGTGTGGCCCTCCAGGGTGACACCGGGTCCCGCAGGGGGCAGCATCAGGCC 441	
Qy	27 -----GluAlaLeuLeuArgValIysSerSerGluArgLeuAlaMetLeuArg 42	
Db	440 CCATCCGGGATCTCCCACCATGCATGTGCAGTCCTCGAGGCCATGGCAATGACGTG 381	
Qy	43 AlaLeuAlaglyMetCysGlyHisArgValLeuProGlyThrGlyala----- 58	
Db	380 TCAPTG-----TCGGGCACCAGCGCATGTCTAGCACAGGGCGTGTGTGGCCACAG 330	
Qy	59 -----SerAlatleAlaAlaThr-----ValThrProIysGlyAlaSerMetIysLeu 74	
Db	329 ACCGTGGGGCATCTTGTGCCACACGTCACAGTCTTGCCAGGGGCAGCACCAGGAAGGCC 270	
Qy	75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgIyslle 94	
Db	269 CCTCCCCCG-----LysThrIlleSerGlnGlu-SerAlaArgVa 108	
Qy	95 ArgGluMetAsn-----LysThrIlleSerGlnGlu-SerAlaArgVa 108	
Db	245 AGGCCACAACCTTAGGGTTGACAGCAGACAGACCCACTGTGCCAGGTGTCTGTGAGACG 186	
Qy	108 laenHisArgLeuproGluGlyHisPro 117	
Db	185 CGCACATCTTCATAGCACTGGTCGGCCT 158	
RESULT 25		
AX237691/c		
LOCUS	AX237691 1659 bp DNA linear PAT 26-SEP-2001	

DEFINITION	Sequence 667 from Patent WO0164886.
ACCESSION	AX237691
VERSION	AX237691.1 GI:15797244
KEYWORDS	
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 Gaiger, A., Algate, P.A. and Mannion, J.
TITLE	Compositions and methods for the detection, diagnosis and therapy of hematological malignancies
JOURNAL	Patent: WO 0164886-A 667 07-SEP-2001;
FEATURES	CORIXA CORPORATION (US)
Source	Location/Qualifiers
	1..1659
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
BASE COUNT	360 a 497 c 511 g 291 t
ORIGIN	
Alignment Scores:	
Pred. No.:	106 Length: 1659
Score:	90.00 Matches: 40
Percent Similarity:	41.54% Conservatives: 14
Best Local Similarity:	30.77% Mismatches: 41
Query Match:	12.71% Indels: 35
DB:	6 Gaps: 6
US-10-087-573-2 (1-141) x AX237691 (1-1659)	
Qy	8 ThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg--- 26
Db	501 ACAGCGTTGGTGGCCCTCCAGGGTGACGAGGGCTCCCGCAGGGCGACCATCAGGCC 442
Qy	27 -----GluAlaLeuLeuArgValIysSerSerGluArgLeuAlaMetLeuArg 42
Db	441 CCATCCGGGATCTCCACACCATGACTGTGCAGTCTTCGGAGCCACTGGCAATGACGTTG 392
Qy	43 AlaLeuAlaGlyMetCysGlyHisargValIysProGlyThrGlyAla----- 58
Db	381 TCATTG-----TGGGGGCACACGCGATGCTAGCACAGGGGCTGTGTGGCCACAG 331
Qy	59 -----SerAlaIleAlaAlaThr-----ValThrProIysGlyAlaSerMetLysLeu 74
Db	330 ACGGTGGGCGATTCTTGCCACACGTCAGTCTTGCCAGGGGCGACACGAGGCGC 271
Qy	75 LysProProArgProGluSerThrIysSerProGluLeuArgGluLeuSerArgLysIle 94
Db	270 CTTCCCGCG-----CTGGCCTCACAGATC 247
Qy	95 ArgGluMetAsn-----LysThrIleSerGlnGluSerAlaArgVa 108
Db	246 AGGCCACAAACTTAGGTGTGACAGCAGACAGAGCCACTGTCCAGGTGTCTGTGAGCG 187
Qy	108 IasnHisArgLeuProGluGlyHisPro 117
Db	186 CGCACATCTTCATAGCACTGGTGGCGCT 159
RESULT 26	
HUMP57B/c	
LOCUS	HUMP57B 1659 bp mRNA linear PRI 09-FEB-1999
DEFINITION	Human mRNA for actin binding protein p57, complete cds.
ACCESSION	D44497
VERSION	D44497.1 GI:927648
KEYWORDS	p57; actin binding protein; WD repeat; leucine zipper.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 Suzuki, K., Nishihata, J., Arai, Y., Honma, N., Yamamoto, K., Irimura, T.

QY	8	ThrAsnPreValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg---	26
Db	501	ACACGCTTGTTGGCCCTCCAGGCTGACAGCGCTCCCGCAGGGGCACCATCAGGCC	442
QY	27	-----GluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg	42
Db	441	CCATCCGGGATCTCCACACCATCACTGTGCAGTCTCCGAGGCACATTCGCAATGACGTTG	382
QY	43	AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla-----	58
Db	381	TCATTG-----TGCGGSCACAGCGCATGCTAGCACAGGGGTGTGTGCCACAG	331
QY	59	-----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu	74
Db	330	ACCGTGGCGCATCTTGTCCACACGTCAGTCTTGCCAGGGGCAGCACAGGANGGCC	271
QY	75	LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle	94
Db	270	CCTCCCCCG-----LysPheSerGlnGluSerAlaArgVal	108

The sequence data in this report represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics -----

Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 207000 bases at least Q40  
Consensus quality: 207365 bases at least Q30  
Consensus quality: 207588 bases at least Q20  
Insert size: 190000; agarose-ff  
Insert size: 207643; sum-of-contigs  
Quality coverage: 9.29x in Q20 bases; agarose-ff

-----  
Quality coverage: 8.50x in Q20 bases; sum-of-contigs  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 24665: contig of 24665 bp in length  
\* 24666 24765: gap of unknown length  
\* 24766 37707: contig of 12942 bp in length  
\* 37708 37807: gap of unknown length  
\* 37808 41993: contig of 4186 bp in length  
\* 41994 42093: gap of unknown length  
\* 42094 51700: contig of 9807 bp in length  
\* 51701 51800: gap of unknown length  
\* 51801 79099: contig of 27299 bp in length  
\* 79100 79199: gap of unknown length  
\* 79200 115446: contig of 36247 bp in length  
\* 115447 115448: gap of unknown length  
\* 115449 150425: contig of 34879 bp in length  
\* 150426 150525: gap of unknown length  
\* 150526 171584: contig of 21059 bp in length  
\* 171585 171685: contig of 20843 bp in length.  
\* 171685 208443: contig of 36759 bp in length.  
Location/Qualifiers  
1..208443  
/organism="Canis familiaris"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9615"  
/clone\_lib="RP81-323GS"  
/clone\_lib="RP81"  
1..24665  
/note="assembly\_fragment"  
clone end:T7  
vector\_side:left  
24766..37707  
/note="assembly\_fragment"  
37808..41993  
/note="assembly\_fragment"  
42094..51700  
/note="assembly\_fragment"  
51801..79099  
/note="assembly\_fragment"  
79200..115446  
/note="assembly\_fragment"  
115447..150425  
/note="assembly\_fragment"  
150526..171584  
/note="assembly\_fragment"  
171585..208443  
/note="assembly\_fragment"  
clone end:SP6  
vector\_side:right  
55095 a 46990 c 47491 g 58067 t 800 others  
ORIGIN

Alignment Scores:  
Pred. No.: 3,136+04 Length: 208443  
Score: 90.00 Matches: 40  
Percent Similarity: 39.46% Conservative: 18  
Best Local Similarity: 27.21% Mismatches: 49  
Query Match: 12.71% Indels: 40  
DB: 2 Gaps: 7

US-10-087-573-2 (1-141) x AC120207 (1-208443)

Qy 2 GluSerThrSerThrThrThrAspPheValAlaGluAsnArgProThrPheGlyGluThr 21

Db 176099 GAACAGACAAACACAAACAAAC-----AAAAATCTCTCTACCTTCGGG---GCG 176052

Qy 22 PheAspValMetArgGluAlaLeuLeuArgValLys-----SerSer 35  
Db 176051 CTTACACTCTCGAAGGGTTGCTGATTTACAGATGGTAGAGCTGAGGTGACGCTCA 175992  
Qy 36 GluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGly 55  
Db 175991 GAGAGGCTAAGCAACCTGACCAAGGTCACGGGCGCTGCGGTTAAAGAATCAGGATTGGG 175932  
Qy 56 -----ThrGlyAlaSerAlaAlaAlaThrValThrProLys----- 68  
Db 175931 ACACAGGACTGTCTGACTCCAGATATGAGGAATGCTTTCTACACACACACGATTACCT 175872  
Qy 69 -----GlyAlaSerMetLysLeuLysProArgProGln----- 80  
Db 175871 CTGGGATCGTGGCTTATCTGGAAGTGTCTCCCCCAACCCCAATCTACACACAAA 175812  
Qy 81 -----SerThrLysSerProGluLeuArgGluLeuSerArgLys 93  
Db 175811 TAAAGGAGACTTGGCAGAGATCCACACCCCGGACTACCTTGGAACCTCTCTCTTGTGCG 175752  
Qy 94 IleArgGluMetAsnLysThrIle-----SerGlnGluSer 105  
Db 175751 GTATATGTGTTTAAACAAGATGTGAGAAATACCTGCCAGGTCTCTTCTTGGCCAGAGCCAA 175692  
Qy 106 AlaArgValAsnHisArgLeu 112  
Db 175691 CGACAGAGCCACACACCGCAATT 175671

RESULT 28  
AB088224/c  
LOCUS 210614 bp DNA linear BCT 11-JUN-2003  
DEFINITION Streptomyces rochei plasmid pSLA2-L DNA, complete sequence.  
ACCESSION AB088224  
VERSION AB088224.1 GI:30698345  
KEYWORDS  
SOURCE Streptomyces rochei  
ORGANISM Streptomyces rochei  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1  
AUTHORS Kinashi,H., Fujii,S., Hatani,A., Kurokawa,T. and Shinkawa,H.  
TITLE Physical mapping of the linear plasmid pSLA2-L and localization of the eryAI and actI homologs  
JOURNAL Biosci. Biotech. Biochem. 62, 1892-1897 (1998)  
REFERENCE 2  
AUTHORS Suwa,M., Sugino,H., Sasaki,A., Mori,E., Fujii,S., Shinkawa,H., Nimi,O. and Kinashi,H.  
TITLE Identification of two polyketide synthase gene clusters on the linear plasmid pSLA2-L in Streptomyces rochei  
JOURNAL Gene 246 (1-2), 123-131 (2000)  
MEDLINE 20231737  
PUBMED 10767533

REFERENCE 3  
AUTHORS Hiratsu,K., Mochizuki,S. and Kinashi,H.  
TITLE Cloning and analysis of the replication origin and the telomeres of the large linear plasmid pSLA2-L in Streptomyces rochei  
JOURNAL Mol. Gen. Genet. 263 (6), 1015-1021 (2000)  
MEDLINE 20408175  
PUBMED 10954087

REFERENCE 4  
AUTHORS Mochizuki,S., Hiratsu,K., Suwa,M., Ishii,T., Sugino,F., Yamada,K. and Kinashi,H.  
TITLE The large linear plasmid pSLA2-L of Streptomyces rochei has an unusually condensed gene organization for secondary metabolism  
JOURNAL Mol. Microbiol. 48 (6), 1501-1510 (2003)  
MEDLINE 22676866  
PUBMED 12791134

REFERENCE 5  
AUTHORS Mochizuki,S., Hiratsu,K. and Kinashi,H.  
TITLE Direct Submission  
JOURNAL Submitted (15-JUL-2002) Haruyasu Kinashi, Hiroshima University, Department of Molecular Biotechnology, Graduate School of Advanced

Sciences of Matter; 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8530, Japan [E-mail: kinashi@hiroshima-u.ac.jp, Tel: 81-824-24-7869, Fax: 81-824-24-7869]

The nucleotide sequence has been determined by using restriction fragments and nested deletion fragments of the ordered cosmid library of pSLA2-L, bp long with G-C content of 72.8 % and carries pSLA2-L is 210,614 bp long with G-C content of 72.8 % and carries 143 ORFs.

Gene prediction was based on the unique codon usage in Streptomyces (Bibb et al., Gene 30:157-66 (1984)) using the FramePlot program of Ishikawa and Hotta (FEMS Microbiol Lett 174: 251-253 (1999)) as implemented at <http://www.nih.go.jp/~jun/cgi-bin/frameplot.pl>. Where possible we chose an initiation codon (atg, ggc, ctg or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5'-13 bp before the initiation codon). If this could not be identified we chose the most upstream initiation codon.

Three-fourths of the pSLA2-L DNA are occupied by biosynthetic genes (three PKS gene clusters for lankacidin, lankamycin and an unknown type II polyketide, a carotenoid biosynthetic gene cluster, and many regulatory genes and others). The range of each biosynthetic gene cluster has been deduced by comparison with similar gene clusters in most cases and may be revised in future.

Location/Qualifiers

1. .210614

/organism="Streptomyces rochei"

/mol\_type="genomic DNA"

/strain="7434AM4"

/db\_xref="taxon:1928"

/plasmid="pSLA2-L"

/notes="linear plasmid"

1. .1992

/notes="left terminal inverted repeat, TIR-L; shows 99.4 % (1981/1992) sequence identity to TIR-R (complement (208623..210614))"

683. 2188

/notes="N-terminal sequence is almost identical (435/437) with that of ORF143 at the right end of pSLA2-L until the inner end of TIR-L"

ORF1 (501 aa)

similar to AL590463 Streptomyces coelicolor putative helicase, SCPI.136 (879 aa); homology is seen until the inner end of TIR-L

/codon\_start=1

/transl\_table=11

/product="putative helicase"

/protein\_id="BAC76459.1"

/db\_xref="GI:30698346"

/translation="MSSTRTQREAAQAEVAVVRALELPVRLAPERGLRTOVIM ATSGSTRVAARSAEKLRGRVLVPSLDLLTQTEARWEAGRTGPMIGVSSLRGED VAFPNITDVEELVDVVRPDKVTATYASLGTLERHRCGLPGWLLIYVDEAHT SGEKLPWAVHNDRLPSRLRYNTATPRLQLQDEAGAGELVASMEDDPOGLFG ACFTLLEADRGICAFQVVCVDITDTQLQAQLLQVEGRSDEVRGARLAALQTA LKASSEENRFTLVFHHVKEAEAFAGLPVAKRLHAAAGFLYPTIWANLCEH KPHRRVLTGFTSGIATDGTVEKGLSVKLVGEGVDTRCEDSVYADVRGSMPL VOAVGRLRIQPCQGVASLVVPLDGTADNMLTSRPYNGALLFRQAPLLTGHG EGFRAARTPGFEVRVLGVGQAQGSAGPASRSITTSAMSSLSISMRAP"

complement (2315..3595)

/notes="ORF2 (426 aa)

similar to AE004736-10 Pseudomonas aeruginosa hypothetical protein (442 aa)

/codon\_start=1

/transl\_table=11

/product="putative secreted protein"

/protein\_id="BAC76460.1"

/db\_xref="GI:30698347"

/translation="MRKXPALLGLPLGLLVAVLAPAPAAHASADTYVGTWAAPTAA PADSTVYEDQLRQTVHUSVAGDSLRVFTNFTFGTSLTIGEVHAAAPAAQGPATV DGTDRVFGKRSATLAPGTQKRSVDPVALPTAGGDLVLSYLPQRTPGSTVSHAA YQHNFAAGDVTGAPDLTSTATGWHFSLGVSVDVRAGTADSVVITLGDSTIDGHT TLDAWRPDLAELRRDGLAGTGVNAG:GGNRLLRDPDPFGSAAGSFAAYFGE SALKRFDVLPQGARAVTVLLGVNDLQPGIAAPASDEVTAEELIAGYQLIERAH EHCLKIYGATITPFAGDTIGYFTPREAVRQVQNDWIRTSAGFDVLDFDAVLDPAR

PDHLLPAYDGGDLHPNDAGMAAARAFPLDLSLR"

complement (3659..4645)

/notes="ORF3 (328 aa)

similar to Y00459-2 Streptomyces griseus regulatory protein, StrR (350 aa)"

/codon\_start=1

/transl\_table=11

/product="StrR-like regulatory protein"

/protein\_id="BAC76461.1"

/db\_xref="GI:30698348"

/translation="MTAALAGNLDQETRSIPSSLPVGYSPRLCGEADAEHVARLDTPD EALPPLVERSTLRVLDGMRHVLAAKAGHTTIEVRLFDGAAEEAFLTAVRSNMTHGL PLQSPERAAQRILACQPHLSRAVAGIAGICAKTVAALRPLAAQPTPQARRGGD GRIRPLDCTGTRKAAELLAQRQASVREVARHAGISPATASVDRRLASGRSPVER LITGARPAAGTGRATPGGAGAEVPIRPDSDPLVRLLRDPSLRSKGRKL LRLQCQAVERTALLAAQVTPPHCTDLVLAELAREYADLWAEFAREVARTDG"

complement (4922..5812)

/notes="ORF4 (296 aa), lankacidin biosynthesis protein similar to AE004625-4 Pseudomonas aeruginosa pyrroloquinoline quinone biosynthesis protein B (304 aa)"

/codon\_start=1

/transl\_table=11

/product="pyrroloquinoline quinone biosynthesis protein B"

/protein\_id="BAC76462.1"

/db\_xref="GI:30698349"

/translation="MILLGTAAGGPPQWNCACALCARRRGELPARSQECVAVSGDG RDWLLNASPDITOLLAAALTEGPGPRTDPVGRVLLTDAEVDHALGLAVRGATGL TVAAPPVGALSAELPVRCGLDRYAPWMDRDTATAGGFVAGGLTVAHPVGTAKK YAPAPDPAPWVCARIEDPATGLVYAFCLATWPDGDDLLASATCALDGTFFSA GELGTSSAGAGQSLMGHLPVAGPGGSLAALARRHRLRIYTHLNNNTLDDPSSAA HAAVREAGVEVLPDGSSELV"

complement (5815..6909)

/notes="ORF5 (364 aa), lankacidin biosynthesis protein similar to AJ277117-5 Gluconobacter oxydans putative pqqE protein (359 aa)"

/codon\_start=1

/transl\_table=11

/product="pyrroloquinoline quinone biosynthesis protein B"

/protein\_id="BAC76463.1"

/db\_xref="GI:30698350"

/translation="MADPAVAGAPAGLIELTHRCPLCHPCVCSNPLELVRREALTCQK WTDLITQARELGVOMHFSGESELPARLDPLVGHARRLGAAYNLVTSVGLTAERAH DLARGVDVQLSLQDAPAAQAIAGARVHTAKLEAARAVTATAGLPLTVLHHRGN IDRTGRMVDLAVLDGADRIELANTQYVGLGRNRAALMPTAAQAAAREAVRHARTY AGGELYVAADYDDRPKFCMDGWSGTQLTVPAGDVLCPAPAYITLTPVENALRR PLGSIWYASRSFNAYRGTMWRPCTCPRHADHGGRCQAFQLTGDAATDPACGL SPHRSLVDAALAEVTDGPVAFVPRGPVPA"

complement (6902..7174)

/notes="ORF6 (90 aa), lankacidin biosynthesis protein similar to AL603642-200 Sinorhizobium meliloti putative Pyrroloquinoline quinone biosynthesis protein D (98 aa)"

/codon\_start=1

/transl\_table=11

/product="pyrroloquinoline quinone biosynthesis protein D"

/protein\_id="BAC76464.1"

/db\_xref="GI:30698351"

/translation="MTGLPEPTVRLRPPGVRLTRDPARGSLALLPERVVVINDTAAAV LAHCDGTSIAGIVERLAEVIEGVSADVEDRELLRLAQRVVDLHG"

complement (7171..7899)

/notes="ORF7 (242 aa), lankacidin biosynthesis protein similar to AE004625-5 Pseudomonas aeruginosa pyrroloquinoline quinone biosynthesis protein C (250 aa)"

/codon\_start=1

/transl\_table=11

/product="pyrroloquinoline quinone biosynthesis protein C"

/protein\_id="BAC76465.1"

/db\_xref="GI:30698352"

/translation="MMSMTREVAAPMSEAEFRQLHALSSSYMDRPFHRRMHEGILL DEBELRWANRWYQRCLPKQDAIIVANCLPEVRRLSRIVYHGDGACAGAEK WRLRLAEAVRRRDEHRLVLAGTFAVDVDFARRRWLAAASGLTSLSPGLL AHELGRLEHYPIAEFGFEYFATREVGFEGRSLDLVVARHVSREQEACVRALA FKCRVLNAVDSLDYHTNGATRS"

complement (7926..8030)

```

/note="ORF8 (34 aa), lankacidin biosynthesis protein
similar to AL603642-197 Sinorhizobium meliloti putative
pyrroloquinoline quinone biosynthesis protein A (31 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein A"
/protein_id="BAC76466.1"
/db_xref="GI:30698353"
/translation="MRTSGKELPAKAWHRPDEFVTIDTGMVTAFFSR"
8420..10057
CDS
/note="ORF9 (545 aa), possible lankacidin resistance
protein"

```

```

Alignment Scores:
Pred. No.: 3.17e+04 Length: 210614
Score: 90.00 Matches: 41
Percent Similarity: 39.74% Conservative: 19
Best Local Similarity: 27.15% Mismatches: 57
Query Match: 12.71% Indels: 34
DB: 1 Gaps: 5

```

US-10-087-573-2 (1-141) x AB088224 (1-210614)

```

QY 5 SerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspVal 24
Db 190198 TCACCCGCGCGCGAGCCCGATCCGATAGCCCAACGACGACGCTACTGCGCGACGCT 190139
QY 25 MetArgGluAlaLeu----- 29
Db 190138 CTAAGAGAGGCGCTTCACGCGAGTGGCGCGCGATCCGCGCGACGCCACTCACT 190079
QY 30 -----LeuArgValIysSerSerGluArgLeuAlaMetLeuArgAlaLeuAlaGlyMet 47
Db 190078 CTCGCGTGGCGGCTCAACGCGACCTGACGCTCTATCGGCTACCGCGGTACCGCGGCGCT 190019
QY 48 CysGlyHisArgValLeuProGlyThrGlyAlaSerAlaIleAlaAlaThrValThrPro 67
Db 190018 TCGGCCACCGC---TGTCGCGAGTGGGACGACAAAGCCCGACGACCCACCAACGCA 189962
QY 68 LysGlyAlaSerMetLysLeuLysProArgProGlnSerThrLysSerProGluLeu 87
Db 189961 CGATCA-----CCGCGATCCACACAGCGCGCTCAGCG 189929
QY 89 ArgGluLeuSerArg--LysIleArgGluMetAsnLysThrIleSerGlnGluSerAla 107
Db 189928 CGAAACCCCGACCGAGAGATCAGACCCAGCAGAGAACCAAGACGAGGAGACCATCG 189869
QY 107 IqValAsnHisArgLeuProGluGlyHisProLeuLeuGlu-----LysArgA 123
Db 189868 TTATCAAC---CTCTGGCCACCGGATGCCCTTACTCAAGCCGACAGCCGCGGAACG 189812
QY 123 laGluTyrPheArgHisLeuArgSerLeu 132
Db 189811 AGAGTTATTTCGCGCGGTCCGGGGCATA 189783

```

```

RESULT 29
LOCUS AC012363/c 217346 bp DNA linear PRI 09-JAN-2002
DEFINITION Homo sapiens BAC clone RP11-438012 from 2, complete sequence.
ACCESSION AC012363
VERSION AC012363.6 GI:15778776
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 217346)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 217346)

```

AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
3 (bases 1 to 217346)  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Ali, J., Kozlowski, A. and Dignan, G.  
The sequence of Homo sapiens BAC clone RP11-438012  
Unpublished (2001)  
3 (bases 1 to 217346)  
Waterston, R.H.  
Direct Submission  
Submitted (25-OCT-1998) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
4 (bases 1 to 217346)  
Waterston, R.H.  
Direct Submission  
Submitted (26-SEP-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
5 (bases 1 to 217346)  
Waterston, R.  
Direct Submission  
Submitted (09-JAN-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Sep 26, 2001 this sequence version replaced gi:15144350.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc  
Contact: sapiens@wustl.wustl.edu  
----- Summary Statistics  
-----  
Center project name: H\_NH0438012  
-----

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.  
McPherson, Department of Genetics, Washington University, St. Louis  
MO. For additional information about the map position of this  
sequence, see http://genome.wustl.edu/gsc

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male  
donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E.,  
Tateno, M., Catanese, J. and de Jong, P.J. (1998) An improved  
approach for construction of bacterial artificial chromosome  
libraries. Genomics 51:1-8. The clone may be obtained either from  
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong  
and coworkers at the Roswell Park Cancer Institute  
(http://bacpac.med.buffalo.edu)

#### VECTOR:

pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-347L11. Actual start of  
this clone is at base position 1 of RP11-438012; actual end is at  
base position 217346 of RP11-438012.

Data from AC079879 was used to finish this clone, AC012363.  
Polymorphisms exist between AC012363 and AC079879. There is a  
single plasmid region from 188387 to 188395.

FEATURES  
source

```

1..217346
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"

```



```

/repeat_region /map="2"
/repeat_region /clone="rp11-438012"
/repeat_region /clone_lib="RPCI-11"
/repeat_region 1..525
/repeat_region /rpt_family="L1"
/repeat_region 545..538
/repeat_region /rpt_family="AT_rich"
/repeat_region 548..1421
/repeat_region /rpt_family="L1"
/repeat_region 212..2522
/repeat_region /rpt_family="Alu"
/repeat_region 3888..4168
/repeat_region /rpt_family="Alu"
/repeat_region 4149..4168
/repeat_region /rpt_family="(A)n"
/repeat_region 4566..4950
/repeat_region /rpt_family="L1"
/repeat_region 4951..5079
/repeat_region /rpt_family="L1"
/repeat_region 5132..5499
/repeat_region /rpt_family="L1"
/repeat_region 5821..6117
/repeat_region /rpt_family="Alu"
/repeat_region 6368..6706
/repeat_region /rpt_family="MER1_type"
/misc_feature 6443..6447
/notes="match to EST T31897 (NID:g613995)"
/repeat_region 6750..6834
/repeat_region /rpt_family="Alu"
/repeat_region 6837..6893
/repeat_region /rpt_family="L2"
/misc_feature 7176..7744
/notes="match to EST AW852630 (NID:g7948147)"
/repeat_region 7378..7505
/repeat_region /rpt_family="L2"
/repeat_region 7699..7977
/repeat_region /rpt_family="Alu"
/misc_feature 8308..8336
/notes="match to EST BF752822 (NID:gl2079498)"
/repeat_region 8398..8680
/repeat_region /rpt_family="Alu"
/repeat_region 8682..8725
/repeat_region /rpt_family="(T)n"
/repeat_region 8691..8921
/repeat_region /rpt_family="Alu"
/repeat_region 8935..9235
/repeat_region /rpt_family="Alu"
/misc_feature 8947..9480
/notes="match to EST BI085105 (NID:gl4503435)"
/misc_feature 9103..9598
/notes="match to EST BI089781 (NID:gl4508111)"
/misc_feature 9735..10199
/notes="match to EST A1684181 (NID:g4895475) tw96d01.x1"
/repeat_region 9921..9949
/repeat_region /rpt_family="AT_rich"
/repeat_region 10024..10362
/repeat_region /rpt_family="MER2_type"
/repeat_region 10412..10719
/repeat_region /rpt_family="Alu"
/repeat_region 10693..10719
/repeat_region /rpt_family="(A)n"
/repeat_region 10831..11014
/repeat_region /rpt_family="MER1_type"
/repeat_region 11186..11459
/repeat_region /rpt_family="Alu"
/misc_feature 11250..11276
/notes="similar to Homo sapiens EST BF769297 (NID:gl2117197)"
/repeat_region 13494..13698
/repeat_region /rpt_family="L2"
/repeat_region 13855..14300
/repeat_region /rpt_family="L1"
/repeat_region 14303..14358

/rpt_family="CT-rich"
14385..14949
/rpt_family="L1"
14950..15506
/rpt_family="MaLR"
15556..15731
/rpt_family="L1"
15738..15788
/rpt_family="MER1_type"
15819..17702
/rpt_family="L1"
17706..17806
/rpt_family="Alu"
17831..18050
/rpt_family="L1"
18094..18211
/rpt_family="MER1_type"
18242..18398
/rpt_family="L1"
18418..18702
/rpt_family="Alu"
18899..19204
/rpt_family="Alu"
19634..19891
/notes="match to EST AI824843 (NID:g5445514) wb02e10.x1"
20004..20582
/notes="match to EST AI769494 (NID:g5236003) wh57f08.x1"
20414..20786
/notes="match to EST AL043595 (NID:g5422982)"
20747..20748
/notes="match to EST U46342 (NID:gl236484)"
21412..21745
/notes="match to EST U46342 (NID:gl236484)"
22152..22728
/notes="match to EST AL042860 (NID:g5935608)"
22164..22441
/notes="match to EST C18353 (NID:gl579955)"
22359..22868

Alignment Scores:
Pred. No.: 3.29e+04 Length: 217346
Score: 90.00 Matches: 36
Percent Similarity: 38.46% Conservative: 9
Best Local Similarity: 30.77% Mismatches: 56
Query Match: 12.71% Indels: 16
DB: 9 Gaps: 4

US-10-087-573-2 (1-141) x AC012363 (1-217346)
QY 16 ProThrPheGlyGluThrPheAspValMetArgGluAlaLeuLeuArgValLysSerSer 35
Db 97847 CCGGCTCGGGCGGACATCTCTCGCGCGCCGCGCCACAGCCGCGCAGCGAGGAGC 97788
QY 36 GluArgLeu-----AlaMetLeuArgAlaLeuAlaGly 46
Db 97787 AAGAGGCTCGGGCGGTGCGGTATCCCACTCCCAAGCCCGCCGCCACCCCGCCGGG 97728
QY 47 MetCysGlyHisArgValLeuProGlyThrGlyAlaSerAlaIleAlaAlaThrValThr 66
Db 97727 CCGCGGGCGCGCGCCCTTACAGGACCACTCCCGCGCTCCGCGCTCCGCGAGCCCGG 97668
QY 67 ProLysGlyAlaSerMetLysLeuLysProProArgProGlnSerThrLysSerProGlu 86
Db 97667 CCC-----GCAGCTCTAAACGCGCGCCCGCCCTCCCGGGCGCGCACCACCGCCGCC 97614
QY 87 LeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAla 106
Db 97613 TCCACCGATTGTTCATTCAGCTGCGGGCTCGCTTCCACC-----CCCCGAGCG 97563
QY 107 ArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAla 123
Db 97562 CGCTCGCGCGTCGCCTA-----GGCAACCCCTCTCTCGAGAGCGCGCC 97518

```

```

RESULT 30
TFEMADF/c
LOCUS      Thiobacillus versutus membrane protein (madF) mRNA, complete cds.
DEFINITION      L36952
ACCESSION      L36952
VERSION      L36952.1 GI:1161286
KEYWORDS      madF gene; membrane protein.
SOURCE      Paracoccus versutus
ORGANISM      Paracoccus versutus
Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
Rhodobacteraceae; Paracoccus.
1 (bases 1 to 1227)
Huitema,F., Duine,J.A. and Canters,G.W.
Sequence of the genes surrounding the genes encoding methyamine
dehydrogenase from Thiobacillus versutus and influence of those
genes on expression of the MADH-subunits in Escherichia coli
Unpublished (1994)
JOURNAL      Original source text: Thiobacillus versutus cDNA to mRNA.
COMMENT      Location/Qualifiers
FEATURES             source
1..1227
    /organism="Paracoccus versutus"
    /mol_type="mRNA"
    /db_xref="taxon:34007"
    <1..1227
    /note="orf 5; protein fragment; putative"
    /codon_start=1
    /transl_table=11
    /protein_id="AAA85385.1"
    /db_xref="GI:1161287"
    /translation="LQVSRDVLISQPHLRDHHAKQLSDWCVSLFAARSP"
    352..360
    /standard_name="Shine-Dalgarno sequence"
    /note="putative"
    367..1191
    /gene="madF"
    367..1191
    /gene="madF"
    /note="madF"
    /codon_start=1
    /transl_table=11
    /product="membrane protein"
    /protein_id="AAA85386.1"
    /db_xref="GI:1161288"
    /translation="WASLDNFDMAAGRTDGVACVAFPPRFTWTRALLIAASAAGG
    AAALMDAHVALVLGLAFAGGLSTWSPGYSISLLRPTGKARGLDMLPTFAT
    HGLGTGALLIGLLGAIGGAGLSGFATSGFLGLLAVIGLAYARGLDFLRVPYQ
    RRAQVPHDARQRFKVVVGGVLSGLDLYTVQTPLLXLTAAVLSGNVAEVAL
    IAFNLGRVLPVNLPLVPTDQIQSWLGRNQERAAIADGAILTAVGAFAFLALA"
BASE COUNT      168 a 406 c 421 g 232 t
ORIGIN
Alignment Scores:
Pred. No.:      81.8      Length:      1227
Score:          89.50      Matches:      32
Percent Similarity: 41.67%      Conservative: 28
Best Local Similarity: 22.22%      Mismatches: 65
Query Match:     12.64%      Indels:      19
DB:              1          Gaps:      4

US-10-087-573-2 (1-141) x TFEMADF (1-1227)

Qy      11 ValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArgGluAlaLeu--- 29
Db      933 GTCCAGTCCCAGCGACGAGCGCGTAAAGCCGCCACGACCCATTTCCGAAAGCGCTCGC 874
Qy      30 -----LeuArgVallysserSerGluArgLeuAlaMet 40
Db      873 CGGTCATGGGGGACTTGGCGCGCGGCTGCGGATAGGCGACGCGAGGAAATCCAGCTG 814
Qy      41 LeuArgAlaLeuAlaGlyMetCysGlyHis-----ArgValLeuProGlyThrGlyAla 58
Db      813 GTGCGCGCCATA---GGCCAGGCGGATCATCGGCGACGCGCGACGCCCGCGAAGCTGTGGC 757
```

```

Qy      59 SerAlaIleAlaIleThrValThrProLysGlyAlaSerMetLysLeuLysProProArg 78
Db      756 GAAGCCGGAAGGCGCGGATCCCGGATCCCGGCGCGCCAGCGGTCGCCAGGATCAGCGC 697
Qy      79 ProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsn 98
Db      696 CCCAGGGCATAGCCAGCCAGCCATGGGTGCGAAAGGTCGGCAGCGACGATCCAGCCGCCG 637
Qy      99 LysThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeu 118
Db      636 CGCCCTTTCCTGCGGCGGCGAGCAACGAGATCGAGTAATAGCCGAGCGCGACCGAGGT 577
Qy      119 LeuGluLysArgLa-----GluTyrPheArgHisLeuArgSerLeu 132
Db      576 CGACAGCGCCCGCCCGCAAGAGCGCGCCAGCCCGACGCGGACGATGCGCCCGGTC 517
Qy      133 LysSerGlnGly 136
Db      516 CATGCCAGCGGC 505

RESULT 31
AC121492/c
LOCUS      Leishmania major chromosome 27 clone LB03590 strain Friedlin,
DEFINITION      complete sequence.
ACCESSION      AC121492
VERSION      AC121492.3 GI:30911132
KEYWORDS      HTG.
SOURCE      Leishmania major
ORGANISM      Leishmania major
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
1 (bases 1 to 96236)
Myler,P.J., Sisk,E., Cawthra,J., Handley,F., Vogt,C., Robertson,L.,
McDonagh,P., Ivens,A., Nguyen,D., Munden,H. and Stuart,K.
Direct Submission
Submitted (18-MAY-2002) Seattle Biomedical Research Institution, 4
REFERENCE      2 (bases 1 to 96236)
AUTHORS      Myler,P.J., Sisk,E., Cawthra,J., Handley,F., Vogt,C., Robertson,L.,
McDonagh,P., Ivens,A., Nguyen,D., Munden,H. and Stuart,K.
JOURNAL      Direct Submission
TITLE      Submitted (22-NOV-2002) Seattle Biomedical Research Institution, 4
REFERENCE      3 (bases 1 to 96236)
AUTHORS      Myler,P.J., Sisk,E., Cawthra,J., Handley,F., Vogt,C., Robertson,L.,
McDonagh,P., Ivens,A., Nguyen,D., Munden,H. and Stuart,K.
JOURNAL      Direct Submission
TITLE      Submitted (20-MAY-2003) Seattle Biomedical Research Institution, 4
COMMENT      Nickerson Street, Seattle, WA 98109-1651, USA
FEATURES             Location/Qualifiers
1..96236
    /organism="Leishmania major"
    /mol_type="genomic DNA"
    /strain="Friedlin"
    /db_xref="taxon:5664"
    /chromosome="27"
    /clone="LB03590"
BASE COUNT      19140 a 28551 c 29412 g 19133 t
ORIGIN
Alignment Scores:
Pred. No.:      1.54e+04      Length:      96236
Score:          89.00      Matches:      39
Percent Similarity: 39.62%      Conservative: 24
Best Local Similarity: 24.53%      Mismatches: 72
Query Match:     12.57%      Indels:      24
DB:              3          Gaps:      6

US-10-087-573-2 (1-141) x AC121492 (1-96236)
```

Qy 4 ThrSerThrThrThrAsnPhValAlaGluAsnArgProThrPheGlyGluThrPheasp 23  
Db 83917 TCTTCTTCATCACCAACTCCGCCAGCTGGCAGCGCGTCTTC---AGCATGTTCTCC 83861  
Qy 24 ValMetArgGluAlaLeuArgVallySerSerGluArgLeuAlaMetLeuArgAla 43  
Db 83860 GCCTTCCTACAGCAGCAGCAGCTGTACCGGGTCTAGTAACCTTCATCATCTCCGCTCT 83801  
Qy 44 Leu-----AlaGlyMetCysGlyHisArgValleu 53  
Db 83800 ACCTCCAGCGGTGGAACACATCGTCTTCATCAGCGGAATATGCTCCAGCAGATATTG 83741  
Qy 54 ProGlyThrGlyAlaSerAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIle 67  
Db 83740 CCGCGCCCTGGGTGAGGTAGTCTCGATCGCTCGATCGCTCGACCAACCTCCGCCACCTCG 83681  
Qy 68 LysGlyAlaSerMetLysLeuLysProProArgProGlnSerThrLysSerProGlu--- 86  
Db 83680 AAGCGCGGTCTATTCAAGCCGATCCCTTCTCCGGAGAAACAACAGCTTCCCAAGGA 83621  
Qy 87 ---LeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSer 105  
Db 83620 TTCTTCGAAGTTTCGCAAGTCAATTGATGAGCGCGCAGCAACACTGTCTGTGCAAA 83561  
Qy 106 AlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGluTyr 125  
Db 83560 TCAGCACTCTTCGCAAGACCAAGGAGGAGCCCTCAACAGAGAAAGTCTCTCA--- 83504  
Qy 126 PheArgHisLeuArg-----SerLeuLysSerGlnGlyValAsnArgLeu 140  
Db 83503 ACGCGAAGGCTACGCGCCAACTCCGCTGCTCTGCGTTTAGCGCAGCATCCAGGCTG 83447  
RESULT 32  
AC084864/C 176562 bp DNA linear PRI 26-MAY-2002  
LOCUS Homo sapiens chromosome 7 clone RP11-738B7, complete sequence.  
AC084864  
ACCESSION AC084864  
VERSION AC084864.4 GI:21217402  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 176562)  
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,  
Saenphimmachak, C., Phelps, K.A., Buckley, D., Kibukawa, M., Raymond, C.  
and Haugen, E.D.  
Direct Submission  
Unpublished  
2 (bases 1 to 176562)  
Kaul, R.K., Olson, M.V., Raymond, C., Clendenning, J. and Haugen, E.  
Direct Submission  
Submitted (23-NOV-2000) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA  
3 (bases 1 to 176562)  
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C.,  
Clendenning, J., Ivey, R.G. and Haugen, E.D.  
Direct Submission  
Submitted (19-APR-2001) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA  
4 (bases 1 to 176562)  
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,  
Saenphimmachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.  
Direct Submission  
Submitted (06-APR-2002) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA  
5 (bases 1 to 176562)  
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,  
Saenphimmachak, C., Phelps, K.A., Buckley, D., Kibukawa, M., Raymond, C.  
and Haugen, E.D.  
Direct Submission  
Submitted (26-MAY-2002) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA

On May 26, 2002 this sequence version replaced gi:2006284.  
----- Genome Center  
Center: University of Washington Genome Center  
Center Code: UWGC  
Web site: http://www.genome.washington.edu  
Contact: uwgctg@u.washington.edu  
----- Project Information  
----- Project name: chr-7  
Center clone name: RP11-738B7 (djs722)  
----- Summary Statistics  
Assembly program: Phrap; version 0.990319  
Consensus quality: 176414 bases at least Q40  
Consensus quality: 176547 bases at least Q30  
Consensus quality: 176562 bases at least Q20  
Insert size: 176562; sum-of-contigs  
Quality coverage: 12.0x in Q20 bases; sum-of-contigs  
-----  
Overlapping Sequences:  
5': RP11-48A19 (UWGC:djs705) AC078946 58194-bp overlap  
3': RP11-437L1 (UWGC:djs723) AC084865 98271-bp overlap  
-----  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
Genbank flat file format but are available as part  
of this entry's ASN.1 file.  
-----  
This sequence was finished as follows unless otherwise noted:  
all regions were either double-stranded or sequenced with an  
alternate chemistry or covered by high quality data (i.e., Phred  
quality >= 30); an attempt was made to resolve all sequencing  
problems, such as compressions and repeats; all regions were  
covered by at least one plasmid subclone or more than one M13  
subclone; and the assembly was confirmed by restriction digest.  
-----  
Sequence Validation:  
This sequence has been validated by Multiple Complete Digest  
fingerprinting. Comparison of the experimentally derived digest  
fragments with sequence-predicted fragments is given below.  
The electronically-digested sequence consists of both insert and  
vector, in order to accurately represent the entire circular BAC.  
Small fragments below a variable cutoff (approximately 400-800 bp)  
are not resolved in the fingerprint and hence do not appear  
in the table. There are no significant remaining discrepancies  
between the experimental and predicted values. Uniquely ordered  
fragments are separated by dashed lines.  
-----  
EcoRI HindIII BglII  
SeqDerMap FngPrnt SeqDerMap FngPrnt SeqDerMap FngPrnt  
-----  
8696 8837 9894 9869 5769 5639  
-----  
<800 6382 6431 2067 2056  
-----  
3060 512 <800 6151 6222  
-----  
12973 12654 449 <800 135 <800  
-----  
1184 22215 21921 17291 16959  
-----  
13704 14723 14460 4160 4079  
-----  
4548 732 772 6476 6481  
-----  
1422 300 <800 89 <800  
-----  
10713 10620 261 <800 2525 2559  
-----

COMMENT

7612	7664	3459	3373	681	<800	183	<800	3024	3237
1478	1494	597	<800	1924	1902	Location/Qualifiers			
1482	1494	2272	2282	3456	3411	1. 176562			
7824	8087	1345	1347	1474	1451	/organism="Homo sapiens"			
1721	1744	2481	2480	2724	2728	/mol_type="genomic DNA"			
313	<800	2806	2809	12250	12362	/db_xref="taxon:9606"			
10245	10620	594	<800	8808	8991	/chromosome="7"			
2957	2902	1091	1115	3945	4079	/clone_lib="RPCI human BAC library 11"			
4167	4235	1129	1115	4649	4766	47363.1.47572			
6118	6170	2198	2282	1854	1902	/standard_name="sWSS3347"			
7425	7499	4177	4278	7735	7953	56730.56808			
1918	2065	24114	24615	19	<800	/standard_name="sWSS3322"			
3930	3930	3744	3797	920	915	121881.122003			
24765	24846	2477	2480	1950	1902	Alignment Scores:			
10168	10620	1254	1273	7997	7953	Pred. No.: 3.48e+04			
543	<800	7662	7831	12518	12362	Score: 88.50			
587	<800	1397	1507	17895	18146	Percent Similarity: 36.92%			
562	<800	3724	3797	2611	2728	Best Local Similarity: 26.15%			
10971	10620	1272	1273	684	<800	Query Match: 12.50%			
4315	4235	3335	3373	2714	2728	DB: 9			
5736	5630	2806	2809	1157	1158	US-10-087-573-2 (1-141) x AC084864 (1-176562)			
748	764	310	<800	5200	5136	Qy 44 LeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAlaIle----- 61			
2861	2902	777	772	1038	1041	Db 70445 CTGGAGGGGGCTGCGAGAGCGCGGGGCGGAGCTGCTCCGTCGGCTCCGCTCCGCTTCTTCA 70386			
10495	10620	9726	9869	2986	2986	Qy 62 -----AlaAlaThrValThrProLysGlyAlaSerMetLysLeu 74			
7	<800	20519	20229	327	<800	Db 70385 GGGTGGCGGCTGGTTCGAGGCTGCTCCGTCGGCGCGAGCGCGCCGCGAGTCTGTG 70326			
		3821	3797	81	<800	Qy 75 LysProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94			
		22	<800	5265	5136	Db 70325 GGGAAAGCGGCTCTCAGACCTCCCGTGCCTCCACTTCGC-----GCCACTCATCC 70272			
		454	<800	3434	3411	Qy 95 ArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeuPro--- 113			
		6414	6431	900	915	Db 70271 CGGGAACCCGCGCAGC-----CGAATGAGACACCGTTCGCCGCC 70230			
		907	915	34	<800	Qy 113 ----- 113			
		2334	2480	470	<800	Db 70229 TCAACTGCCCATTCCTGTTAATAGCGCTAGTGGCACCCAGGTTCCAGCGCTTCCCCAA 70170			
		284	<800	3878	3862	Qy 114 -----GlutGlyHisProLeuLeuGluLysArgAlaGluTyrPheArgHisLeu 129			
		3580	3509	3274	3411	Db 70169 GATCGCGGCTGGACGGCTTGCCTCTGGCGCTCACCAGTCA-----GGCATCCC 70119			
		150	<800	8043	7953	Qy 130 ArgSerLeuLysSerGlnGlyValAsnArg 139			
		1262	1273	1863	1902	Db 70118 AGGACTATGGCAGTGAGGTGTCAACCCGG 70089			
		5121	5107	2821	2986	RESULT 33			
						AE001863			
						LOCUS			
						DEFINITION			
						ACCESSION			
						VERSION			
						KEYWORDS			
						SOURCE			
						ORGANISM			
						REFERENCE			
						AUTHORS			
						TITLE			

Genome sequence of the radioresistant bacterium Deinococcus

White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.J., Nelson, W.C., Richardson, D.L., McFarr, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zaleski, C., Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M. et al.

198677 bp DNA linear BCT 22-NOV-1999  
Deinococcus radiodurans R1 section 2 of 2 of the complete  
chromosome 2.  
AE001863.1 GI:6460670  
Deinococcus radiodurans  
Deinococcus radiodurans  
Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;  
Deinococcaceae; Deinococcus.  
1 (bases 1 to 198677)  
White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.J., Nelson, W.C., Richardson, D.L., McFarr, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zaleski, C., Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M. et al.

JOURNAL MEDLINE PUBMED	radiodurans R1 Science 286 (5444), 1571-1577 (1999) 20036896 10567266
REFERENCE	2 (bases 1 to 138677)
AUTHORS	White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Moffat, K.S., Jin, J., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Uitterback, T., Zalewski, C., Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W., Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S., Smith, H.O., Venter, J.C. and Fraser, C.M.
TITLE	Direct Submission
JOURNAL	Submitted (02-NOV-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES	Location/Qualifiers
source	1. .138677
gene	/organism="Deinococcus radiodurans"
	/mol_type="genomic DNA"
	/strain="R1"
	/db_xref="taxon:1299"
	/chromosomes="2"
	/complement (91. .1926)
	/gene="DRA0206"
CDS	/gene="DRA0206"
	/note="similar to GB:AE000520 percent identity: 43.92; identified by sequence similarity; putative"
	/codon_start=1
	/transl_table=11
	/product="oligodeoxyphosphatase F, putative"
	/protein_id="AAF12423.1"
	/db_xref="GI:6460718"
	/translation="MLARACPLRGAGRCRSPAPQSCPGOMERERWARLPMTTQ SNLEAVEHLMVPADDAQWASVAPFALQADLTAGDVPVWLERNSLSARLOGSGS KLTSHADLHDDAQICQRTQIVAEVLPQERASQALTEKLLAVPGVPGDFALNLR RFQDAALFRENVDLQTHNQINSHSVITGNQVQLGGQELTLPQAKQSGDPAAR QEAAWALSANAAIAPELDALMDLITREQLAHNADDEANFRDVRKELDRVDYTP ECLDFHARSDEVVPLSLGVLDIARQLGSLRFPWYNRNLLDPGERSLRPFAT GADLEAAQOVYESIDPLAAFRQWRQRESGLLDESPPQKTHAYCNYFPPTNEPFL MNVQPAEVRVLEHVEGHAFHGYSGDRQPLVWNRSPLEFEVPEPSMAEFLTLDEL SHVFSDELAERVKQLEGVIAFLPWAQMDAFQHWLYAEAPEDVTIADLDAKWELE RTHFPIDNSGLDESVRAGWYHYHIFQVFFYIYAMCYLAAVINRAARQDPAAAL GRKDALRGNVTSVPELFRAGAEFRDFEYIRGLMGLKEQLG"
gene	complement (2005. .3645)
CDS	/gene="DRA0207"
	/note="similar to PID:1653945 percent identity: 49.28; identified by sequence similarity; putative"
	/codon_start=1
	/transl_table=11
	/product="conserved hypothetical protein"
	/protein_id="AAF12422.1"
	/db_xref="GI:6460717"
	/translation="MNFEVLSPRLSSVAHNLNRSALKEEMSKRSLSALLGLALL SAGSPENGPOAPQLNPNTTEAPQTPSGKVRGLWDAFGPGIKTPEEVSLVTAR ANVNTLYVQVGERGCYCNAAMPETDDPEVPAGDPDLADVIAXAHAKDIEVNAWII TALNWSVTPPKSAEHAFTNGHPNAAGRNWLTIVKADGTVKAGADWLDPGDAAAY IRNMYVSARKVIMFEDRVRYPDYNPGVPGVQWYNETALARIETATGATPDPA ADQWSARRRQQTNLVRETLAKAVKPGVAIAAATITYGEGPADEAAFRSPYAE VQGMILTWKSGVLDQNMVNYKRFVPLQAGNFQDMTFAAGLRQRYPHTEQVSGAA IYLANDASTVQIRKTAQAAGLWAGYSYRTPKQDVNNEGKRGAAEVLADLKLRLTDAGG PFTTFVKHAPLPAASVRLSGELRAARPGHSHSVELVRAGGTWNDVVAATFSDGQRYG FLNLPGQAFDVRVDGKILLPNVQATPGRTVMLEPITLN"
gene	complement (3751. .4860)
CDS	/gene="DRA0208"
	/complement (3751. .4860)
	/note="similar to PID:1001216 PID:1001269 percent identity: 66.97; identified by sequence similarity; putative"
	/codon_start=1
	/transl_table=11
	/product="conserved hypothetical protein"
	/protein_id="AAF12421.1"
	/db_xref="GI:6460716"
	/translation="MTGSETRTPLOMAWRRYKSRPGVIAGWVLIALLVIALSPFL APNVTQHEGYEQRPQVHLHQELHRPVGYSTTKEDPVTFAKKTVPDHTAHLPL VQFVKGEPYFGLGSDRHLFGVPGVQDEYGEKVRPRFFLGTGDLGRLFSFTLVG AQVMTGVVVVGLTISFALGILGVSQYFGWVDTLTIQRVVEVLLSPFRLPILLAT LIAPRFTSWVGTIVAVLALIGWASLVRQGVLAARNLDYVSAQAQWAGNSLRVI LRHITNLSFLVLTATLALPGVILGSLTSLFLGLIKEPMTSWGILLDRASKLEVVS SYEWLLPGLVFVSVLAFNFMGDALRDAADTQSR"
gene	complement (4857. .5843)
CDS	/gene="DRA0209"
	/complement (4857. .5843)
	/note="similar to PID:1001492 PID:1001545 percent identity: 62.93; identified by sequence similarity; putative"
	/codon_start=1
	/transl_table=11
	/product="peptide ABC transporter, permease protein"
	/protein_id="AAF12420.1"
	/db_xref="GI:6460715"
	/translation="MLCCHMLNLTLLRILGMIFTLVLSVLCFLVKLPQGSFIDQVL EDPRFSKETVANITRQLGDPQWAKQYLNWITGVVTRLDYFSGFASGAPVTSIGERL GWVFLAALTALTWTLIAVPLGIYTFNRYGIPAQIANFLGYVGLAIPDPLAALLLVY LTRVQLGNTVGGIIFRPMIDAPWSFAKVLDDLALHLPVLAVGLEGVAGVLRQWRAS TLDVLSQDIYRTAKAGLPQGRVWKHAVNAINPLISLAGLSLPTISGTLIIISVMS LPTIGPLLYDSLUNKQOYTANTLMLSAFLLIIONLLSLLALANADPRVYS"
gene	complement (5964. .7804)
CDS	/gene="DRA0210"
	/note="This region contains an authentic frame shift and is not the result of a sequencing artifact; similar to GB:AE000657 percent identity: 51.19; identified by sequence similarity; putative"
	/codon_start=1
	/transl_table=11
	/product="transcriptional regulator, GntR family"
	/protein_id="AAF12419.1"
	/db_xref="GI:6460714"
	/translation="MWSNNKTFSELVKRYGCAVSSPAPLPPEPPAWWALDPADAAP AVVLAQGLRRRTIESGELSAGTALPAERELAAKLGVSRTLROGLLEAGLRRRR GSQTFVAGTPTHTSPQLGLSSFSDEVAQGVPGARVLSFTQGRPTQEALESLGLTAA EQVYLRRLTSDGPLAVEDSLPAALIGPLTAQDVTDAISLYGLLQORHLAPARATR HLRAQNASAEHLALHLLGVPLGAULTTERTVSWLASGRVPEYARACYGRDIFVMDLKG EDH"
gene	8603. .9806
CDS	/gene="DRA0212"
	/note="This region contains an authentic frame shift and is not the result of a sequencing artifact; similar to GB:U42023 SP:P44861 PID:1005712 PID:1220844 PID:1205001 percent identity: 56.04; identified by sequence similarity; putative"
	/codon_start=1
	/transl_table=11
	/product="conserved hypothetical protein"
	/protein_id="AAF12418.1"
	/db_xref="GI:6460713"
	/translation="WTDPRTEQVHPDYADLDTLAPDLALAAADLQGLAVRAVQAAA PQLTAALNAAVPQLERGLVYVGAGTSGRLGVLDATELTFTFSWPERAVPLIAGE

RAIRQAVEGADEDEAAGRDVQAVNIGPDDVLIAVAASGTTPTVILGAARSRAAGALT  
VGLANPGAPLIAADPCVLLDTPETISGSTRKAGTAQKIALNTLSALMVLKGL  
YGNLAVDVATNAKLEDRRLVQVHATGADADAQAALSECAGSVKTLAVMLKGLGA  
QQAARLEGAGGHARQVLGEAGLTSAS"  
complement (10796..11398)  
/gene="DRA0214"  
complement (10796..11398)  
/gene="DRA0214"  
/note="similar to GB:M99166 SP:P30849 PID:148264 GB:U00096  
PID:1651499 percent identity: 54.36; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="trp repressor binding protein wrbA, putative"  
/protein\_id="AA12417.1"  
/db\_xref="GI:6460712"  
/translation="MTAPVKLAIVFSTGTGYAVAQEAAGRAAGAEVRLKVVRET  
APQVIDGDQWANKIANEAKDVEATPADLEWAEIAIVFSSPTRFGGATSQMRAFDITL  
GLWSSGLKANTFSAITSAONVNGOETTLQTLVMTAMHGAULTPPGYTDEVIFKS  
GNNPYGASVTANGQPLLENDRAISRQVRQVELTAKULG"  
11664..12059  
/gene="DRA0215"  
11664..12059  
/gene="DRA0215"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA12499.1"  
/db\_xref="GI:6460794"  
/translation="MGGLHPRQLGHLRSLDELPMFGWLNPREVAFLKAGLNAAG  
LTGMGTPHPALGPPAFAALLSDNEQARLGLDGLVRLVAHHLTCLTLGAPLLDND  
TERVQRIEATRPPLFVDEVMAGREDEGA"  
12068..13069  
/gene="DRA0216"  
12068..13069  
/gene="DRA0216"

Alignment Scores:  
Pred. No.: 4e+04 Length: 198677  
Score: 88.50 Matches: 37  
Percent Similarity: 49.59% Conservative: 24  
Best Local Similarity: 30.08% Mismatches: 36  
Query Match: 12.50% Indels: 26  
DB: 1 Gaps: 8

US-10-087-573-2 (1-141) x AE001863 (1-198677)

QY 25 MetArgGluAlaLeuLeu-----ArgValLysSerSerGluArgLeuAla 39  
DB 147384 CTGCGTAGCGGTGCTGAGCTGCATGCGCGCGCATCGAAGAACCGGCTCAGAGTCT 147443

QY 40 MetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGly----- 57  
DB 147444 -----CAGGCCCTG-----CGCGGATTCTCGCGGAGAGGACGGTG 147482

QY 58 -----AlaSerAlaIleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLys 75  
DB 147483 ACGTCAAGCGGCGCGCGCGCTGCGCGCTTCACAGCCCAATCTCAGCGGTGCGT 147542

QY 76 Pro-----ProArgProGlnSerThrLysSerProGluLeu-----ArgGluLeuSer 91  
DB 147543 CCAAGTCCAGCGGCTACTCCGATCGCATCGCCCAACCGCAACCGGACCGGAGTGGA 147602

QY 92 ArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArg 111  
DB 147603 AAACGGCTCGGAGCTGGAAAGAAAGTCCGCAAGGAGCGCGGAGGAGGAGTCCCGGCGC 147662

QY 112 LeuPro-GluGlyHisProLeu-----LeuGluLysArgAlaGluTyrPheAr 127  
DB 147663 ---CCTGCACCGGCACCGCTCCCATTCCTCCATCCCTGCGCAGCGCGCCCTGCG 147719

QY 127 gHisLeu 129

DB 147720 CCCCCTC 147726

RESULT 34

LOCUS BD180073/c 846 bp DNA linear PAT 15-MAY-2003

DEFINITION Highly thermophilic bacterium-derived protein and gene encoding it.

ACCESSION BD180073

VERSION BD180073.1 GI:30790991

KEYWORDS JP 2002325574-A/564.

SOURCE Thermus thermophilus

ORGANISM Thermus thermophilus

Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae; Thermus.

REFERENCE 1 (Bases 1 to 846)

AUTHORS Kuramitsu, N. and Yokoyama, S.

TITLE Highly thermophilic bacterium-derived protein and gene encoding it

JOURNAL Patent: JP 2002325574-A 564 12-NOV-2002;

THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH

OS Thermus thermophilus

COMMENT PN JP 2002325574-A/564

PF 12-NOV-2002

PF 23-FEB-2001 JP 2001116171

PI NARUKI KURAMITSU, SHIGEYUKI YOKOYAMA

PC C12N15/09, C12N15/09, C07K14/195, C12N1/15, C12N1/19, C12N1/21, PC C12N5/10,

PC C12N9/88, C12P21/02, (C12N9/88, C12R1:01), (C12N15/09, C12R1:01), (C12P21/02, C12R1:01), (C12N15/00, C12N5/00, C12N15/00, C12R1:01)

CC Highly thermophilic bacterium-derived protein and gene CC encoding it

FEATURES

source

1..846 Location/Qualifiers

/organism="Thermus thermophilus"

/mol\_type="genomic DNA"

/db\_xref="taxon:274"

BASE COUNT 89 a 324 c 283 g 150 t

ORIGIN

Alignment Scores:

Pred. No.: 71.4 Length: 846

Score: 88.00 Matches: 34

Percent Similarity: 49.18% Conservative: 26

Best Local Similarity: 27.87% Mismatches: 48

Query Match: 12.43% Indels: 14

DB: 6 Gaps: 6

US-10-087-573-2 (1-141) x BD180073 (1-846)

QY 25 MetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArgAlaLeu 44  
DB 398 CTTTCCAGGGCGGCTCCCGCACCTCTCCGCGAGAGGTCTTCGGTCTCCGCTCAGC 339

QY 45 AlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSer-----AlaIleAla 62  
DB 338 -----ARGCGGAAAGCGGCTTCTCCCGCTCAGAGGCGCTTCCAGGCGCTTCTCC 285

QY 63 AlaThrValThrProLysGlyAlaSerMetLysLeuLys-----ProArgPro 79  
DB 284 ACCCAAAGACCTCCAAAGAGGAGCATCTCAGGCTGAAGAGGGCTTCAGGAGGAGGCC 225

QY 80 GlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLys 99  
DB 224 AGGACCAACC---CACCCCGGAGGCGGCTCAGGAGGAGGTCCAGGAGAAAGCGCGGC 168

QY 100 ThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisPro----- 117  
DB 167 AGGGGAAGAGGAGGCGGCGGCAAGAGCCAGTAGAGCGCGCGGAGAAACCCCGGACC 108

QY 118 LeuLeuGluLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyVal 137  
DB 118 LeuLeuGluLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyVal 137

```

Db      107 CTCTCCAGCCCGCAGCGAGGTAC-----CGCCCCATCCAGACCACGGGGTGG 60
QY      138 AsnArg 139
       : : : :
Db      59 AGCGG 54

RESULT 35
AP005426
LOCUS   Oryza sativa (japonica cultivar-group) DNA linear HTG 13-JUN-2002
DEFINITION
*** SEQUENCING IN PROGRESS ***
ACCESSION
AP005426.1 GI:21396536
VERSION
HTG; HTGS PHASE2.
KEYWORDS
Oryza sativa (japonica cultivar-group)
SOURCE
Oryza sativa (japonica cultivar-group)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
1 Sasaki,T., Matsumoto,T. and Katayose,Y.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 9, PAC
Clone:P0668D04
JOURNAL
Published Only in Database (2002)
REFERENCE
2 (bases 1 to 148797)
AUTHORS
Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE
Direct Submission
JOURNAL
Submitted (12-JUN-2002) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasak@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..148797
   /organism="Oryza sativa (japonica cultivar-group)"
   /mol_type="genomic DNA"
   /cultivar="Nipponbare"
   /db_xref="taxon:39947"
   /chromosome="9"
   /clone="P0668D04"
BASE COUNT 42512 a 32195 c 31612 g 42428 t 50 others
ORIGIN

Alignment Scores:
Pred. No.: 3.14e+04 Length: 148797
Score: 88.00 Matches: 39
Percent Similarity: 39.16% Conservative: 26
Best Local Similarity: 23.49% Mismatches: 63
Query Match: 12.43% Indels: 38
DB: 2 Gaps: 5

US-10-087-573-2 (1-141) x AP005426 (1-148797)
QY      2 GluSerThrSerThrThrThrAsn-----PheValalaGluAsnArgProThrPheGly 19
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      145582 GAGAAACACGACGACGCGGAAGTACACTGACACTATAATACACAGTAGACT----- 145632
QY      20 GluThrPheAspValMetArgGluAlaLeuLeuArgVallySerSerGluArgLeuAla 39
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      145633 -----TGTCAGCGAGGAGGCGATGCATGGAAGCTCTAGCACAGGTAGTATCGA 145680
QY      40 MetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSer 59
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db      145681 GCATGCATGAGCTTACCGGAGATGGAGTTTTCAGATCTTTAAAGCGGAGATAGCTATG 145740
QY      60 AlaileAlaIaIaThrValThrProLyseGlyAlaSerMet-----Lys 73
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      145741 GCGGTAGGTGCAGATGTCTTCCTCTCGACGACCCACTCTCTCTCTCCCAA 145800
QY      74 LeuLysProProArgProGlnSer-----ThrLysSerProGluLeuArgGluLeuSer 91
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      145801 ATTAAACCCACACATCACCATCAACGAGATCGATCGATCGATGAGGAGGAGAAAGC 145860
QY      92 ArgLysIleAeGluMetAsnLysThrIleSer----- 102
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      145861 AAGAAGCGAAGAAAGAAAGAACAGCTCAAAATTTGCGAGATTTTCGGCTTGGTTTA 145920
QY      103 -----GlnGluSerAlaAeGValAsnHis 110
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      145921 CATCTCTAAAGAAAGGAAAGAGAACGAGAGGAAAGCTAGCTGGAGAGCAAGTGAG 145980
QY      111 ArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGluTyrPheAeGHisLeuArg 130
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      145981 AGAGAGAGAGAGGATCGAAGATGAAGAGGAGGAGGAGCTAGCAAGCATGAAGG 146040
QY      131 SerLeuLysSerGlnGly 136
       : : : : : : : : : :
Db      146041 ACTAGTCAAAACAGTGGT 146058

RESULT 36
AP005429/c
LOCUS   Oryza sativa (japonica cultivar-group) DNA linear HTG 13-JUN-2002
DEFINITION
*** SEQUENCING IN PROGRESS ***
ACCESSION
AP005429.1 GI:21396539
VERSION
HTG; HTGS PHASE2.
KEYWORDS
Oryza sativa (japonica cultivar-group)
SOURCE
Oryza sativa (japonica cultivar-group)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
1 Sasaki,T., Matsumoto,T. and Katayose,Y.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 9, PAC
Clone:P0701F11
JOURNAL
Published Only in Database (2002)
REFERENCE
2 (bases 1 to 156242)
AUTHORS
Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE
Direct Submission
JOURNAL
Submitted (12-JUN-2002) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasak@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..156242
   /organism="Oryza sativa (japonica cultivar-group)"
   /mol_type="genomic DNA"
   /cultivar="Nipponbare"
   /db_xref="taxon:39947"
   /chromosome="9"
   /clone="P0701F11"
BASE COUNT 41848 a 34933 c 34830 g 44481 t 150 others
ORIGIN

```





/product="putative transcriptional regulator"

Db	261528	GGCAAGCTCGACGGCATCGCATGCGGCTC-----CCGCTCCACCGCGGTGCGCCACC	261477
QY	85	ProGluLeuArgGluLeuSerArgLysLysleArg-----GluMetAsnLysThrIleSer	102
Db	261474	GACCTGGTCTCGAGCTCCAGCGGAGGTCCACCAAGGACGAGGTCAACGC-CGCGTT---	261419
QY	103	GInGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLys---	121
Db	261418	CAAGAAGGCGCGCGA-----CGACGGCGACCTCAAGGGCATCTCTGTCTACACCGGGA	261365
QY	122	-----ArgAlaGluTyrPheArgHisLeuArgSerLeuLys	133
Db	261364	CGCGATGCTCTCTCGGACATCACCGGACCGCGCTCTTCGACCTTCGACTCTCTCCCT	261305
QY	134	SerGIngly 136	
Db	261304	GACCATGGT 261296	
RESULT 39			
AC125900			
LOCUS			
DEFINITION		300813 bp DNA linear HTG 15-NOV-2002	
		Rattus norvegicus clone CH230-319G17, *** SEQUENCING IN PROGRESS	
		***, 12 unordered pieces.	
ACCESSION		AC125900	
VERSION		AC125900.3 GI:25008624	
HTG, HTGS, PHASE1, HTGS, DRAFT, HTGS, ENRICHED.			
KEYWORDS		Rattus norvegicus (Norway rat)	
SOURCE		Rattus norvegicus	
ORGANISM		Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
REFERENCE		1 (bases 1 to 300813)	
AUTHORS		Muzny, D.M., Metker, M., Lee, M., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flag, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, B., Geis, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Haviak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, Y., London, P., Longacre, S., Lopez, J., Loruschwala, L., Louissegh, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenan, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundaasa, W., Murphy, M., Nair, D., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemeh, O., Okwono, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Pu, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shauman, S., Shen, H., Shetty, J., Sivatsreyn, A., Sisson, I., Sitter, C.D., Smales, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,	

Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 300813)  
Worley, K.C.

Direct Submission  
Submitted (02-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 300813)  
Rat Genome Sequencing Consortium.

Direct Submission  
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 15, 2002 this sequence version replaced gi:23908040. and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information  
Center project name: GNMQ  
Center clone name: CH230-319G17  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 246252 bases at least Q40  
Consensus quality: 251299 bases at least Q30  
Consensus quality: 254358 bases at least Q20  
Estimated insert size: 292865; sum-of-contigs estimation  
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html))  
\* NOTE: This sequence may represent more than one 'clone'.  
\* NOTE: This is a 'working draft' sequence. It currently consists of 12 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 93046: contig of 93046 bp in length  
\* 93047 93146: gap of unknown length  
\* 93147 104391: contig of 11245 bp in length  
\* 104392 104391: gap of unknown length  
\* 104492 109137: contig of 4646 bp in length  
\* 109138 109237: gap of unknown length  
\* 109238 116031: contig of 6794 bp in length  
\* 116032 116131: gap of unknown length  
\* 116132 136149: contig of 20018 bp in length  
\* 136150 136249: gap of unknown length

\* 136250 148163: contig of 11914 bp in length  
\* 148164 148263: gap of unknown length  
\* 148264 151741: contig of 3478 bp in length  
\* 151742 151841: gap of unknown length  
\* 151842 198973: contig of 47132 bp in length  
\* 198974 199073: gap of unknown length  
\* 199074 272301: contig of 73228 bp in length  
\* 272302 272401: gap of unknown length  
\* 272402 274687: contig of 2286 bp in length  
\* 274688 274787: gap of unknown length  
\* 274788 276007: contig of 1220 bp in length  
\* 276008 276107: gap of unknown length  
\* 276108 300813: contig of 24706 bp in length.

FEATURES  
Location/Qualifiers  
1..300813  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-319G17"  
26866..27777  
/note="clone boundary  
clone\_end:Sp5  
site:  
end\_sequence:BZ143610"  
93147..96538  
/note="wgs contig"  
100544..102491  
/note="wgs contig"  
102915..104391  
/note="wgs contig"  
108031..109137  
/note="wgs contig"  
113568..116031  
/note="wgs contig"  
122298..123933  
/note="wgs contig"  
123984..126487  
/note="wgs contig"  
128538..133076  
/note="wgs contig"  
133141..134325  
/note="wgs contig"  
135058..136149  
/note="wgs contig"  
138250..138009  
/note="wgs contig"  
138762..141073  
/note="wgs contig"  
150640..151741  
/note="wgs contig"

BASE COUNT 72134 a 55547 c 57928 g 70085 t 45119 others  
ORIGIN

Alignment Scores:  
Pred. NO.: 7.16e+04  
Score: 88.00  
Percent Similarity: 42.37%  
Best Local Similarity: 27.97%  
Query Match: 12.43%  
DB: 2  
Gaps: 4

US-10-087-573-2 (1-141) x AC125900 (1-300813)

Qy 24 ValMetArgGluAlaLeuLeuArgValLysSerGluArgLeuAlaMetLeuArgAla 43  
Db 160087 GTAATAAAATACCTCTAATAAAGAAATATAATAGGAAATGCACTTATTGCTTC 160146  
Qy 44 LeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAlaLeuAla 63  
Db 160147 CTACAA-----CAGCAGGCTCTCCCGGGTATGGTGCACGCCGACGACAG 160194  
Qy 64 ThrValThrPro-----LysGlyAlaSerMetLysLeu 74

Db 160195 CAAGTGCCTCCGTTACAGTCCCATTCCTAGCCTTGATCTGCTGCTTTAGATTA 160254  
 Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
 Db 160255 GAACCAAGGGAGTGAAGCGCTG-----CCAGTCTTGGCACCTTG----- 160296  
 Qy 95 ArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGlu 114  
 Db 160297 -----CGGTGTGATCTCTCCACCGTCAAGTCTACTCTACTATGGCCAGAG 160341  
 Qy 115 GlyHisProLeuLeuGluLysArgAlaGluTyPheArgHisLeuArgSerLeu 132  
 Db 160342 GGTATCTCAAAGCAGCAAGAAAGAACCACTGTGCTGCTCATCTCAGGCCACTA 160395  
 RESULT 40  
 AC118450  
 LOCUS Rattus norvegicus clone CH230-144P6, \*\*\* SEQUENCING IN PROGRESS  
 DEFINITION \*\*\* 35 unordered pieces.  
 AC118450 308821 bp DNA linear HTG 12-OCT-2002  
 AC118450.3 GI:23908096  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
 KEYWORDS Rattus norvegicus (Norway rat)  
 SOURCE Rattus norvegicus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 308821)  
 Muzny,D.,Maric,E., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,  
 Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,  
 Anyalebechi,V., Ayagi,A., Ayodeji,M., Baca,E., Baden,H.,  
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
 Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,  
 Cardenas,V., Carter,K., Cavazos,I., Ceasat,H., Center,A.,  
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,  
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,  
 Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,  
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,  
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,  
 Fernandez,S., Finley,M., Flaggs,N., Forbes,L., Foster,M., Foster,P.,  
 Fraser,C.M., Gabisi,A., Gante,R., Garcia,A., Garner,T., Garza,M.,  
 Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guervara,W.,  
 Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,  
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,  
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,  
 Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,  
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
 Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
 Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,B., Li,Z., Liu,J.,  
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
 Lorensuhewa,L., Louissegh,H., Lozado,R.J., Lu,X., Ma,J.,  
 Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,  
 Mangum,B., Mapua,P., Martin,M., Mahmoud,M., Malloy,K., Martinez,E.,  
 Mahoney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,  
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
 Nwakoleneh,O., Okwuonu,G., Olarnpusagoon,A., Pal,S., Parks,K.,  
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,  
 Plopper,F., Polndexter,A., Popovic,D., Primus,E., Pu,L.-L.,  
 Puzio,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
 Rives,C., Rodkey,I., Rojars,A., Rose,M., Rose,R., Ruiz,S.J.,  
 Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,  
 Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,  
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,  
 Valas,R., Vera V. Villanana,D., Waldron,L., Walker,B., Wang,J.,  
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
 Williams,G., Willson,R., Wleczky,K., Wooden,H., Worley,K.,

Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
 Weinstock,G. and Gibbs,R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 308821)  
 Worley,K.C.  
 Direct Submission  
 Submitted (17-APR-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 308821)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (12-OCT-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Oct 12, 2002 this sequence version replaced gi:21903188.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GWHK  
 Center clone name: CH230-144P6  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 0 bases at least Q40  
 Consensus quality: 0 bases at least Q30  
 Consensus quality: 0 bases at least Q20  
 Estimated insert size: 268894; sum-of-contigs estimation  
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html))  
 \* NOTE: This sequence may represent more than one clone.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 35 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 1022: contig of 1022 bp in length  
 \* 1023 1122: gap of unknown length  
 \* 1123 2637: contig of 1515 bp in length  
 \* 2638 2737: gap of unknown length  
 \* 2738 4675: contig of 1937 bp in length  
 \* 4675 4774: gap of unknown length  
 \* 4775 6221: contig of 1447 bp in length  
 \* 6222 7397: gap of unknown length  
 \* 7398 7497: gap of unknown length  
 \* 7498 8902: contig of 1305 bp in length  
 \* 8903 13598: contig of 4696 bp in length  
 \* 13599 13698: gap of unknown length  
 \* 13699 15957: contig of 2259 bp in length

\* 15958 16057: gap of unknown length  
\* 16058 21808: contig of 5751 bp in length  
\* 21809 21808: gap of unknown length  
\* 21909 21809: contig of 1322 bp in length  
\* 22331 21909: gap of unknown length  
\* 22331 22331: contig of 1352 bp in length  
\* 22331 24682: contig of 1352 bp in length  
\* 24683 24782: gap of unknown length  
\* 24783 28502: contig of 3720 bp in length  
\* 28503 28502: gap of unknown length  
\* 102851 102850: contig of 74248 bp in length  
\* 102951 102950: gap of unknown length  
\* 196281 196281: contig of 93331 bp in length  
\* 196282 196381: gap of unknown length  
\* 196382 197669: contig of 1288 bp in length  
\* 197670 197669: gap of unknown length  
\* 197770 199047: contig of 1278 bp in length  
\* 199048 199147: gap of unknown length  
\* 200429 200429: contig of 1282 bp in length  
\* 200430 200529: gap of unknown length  
\* 200530 234122: contig of 33593 bp in length  
\* 234123 234222: gap of unknown length  
\* 234223 234931: contig of 1269 bp in length  
\* 234932 235591: gap of unknown length  
\* 235592 238938: contig of 3347 bp in length  
\* 238939 239038: gap of unknown length  
\* 239039 240391: contig of 1353 bp in length  
\* 240392 240491: gap of unknown length  
\* 240492 243023: contig of 2532 bp in length  
\* 243024 243123: gap of unknown length  
\* 243124 244444: contig of 1321 bp in length  
\* 244445 244544: gap of unknown length  
\* 244545 248267: contig of 3723 bp in length  
\* 248268 248367: gap of unknown length  
\* 248368 249563: contig of 1196 bp in length  
\* 249564 249663: gap of unknown length  
\* 249664 251245: contig of 1582 bp in length  
\* 251246 251345: gap of unknown length  
\* 251346 253225: contig of 1880 bp in length  
\* 253226 253325: gap of unknown length  
\* 253326 257369: contig of 4044 bp in length  
\* 257370 257469: gap of unknown length  
\* 257470 259280: contig of 1811 bp in length  
\* 259281 259380: gap of unknown length  
\* 259381 261051: contig of 1671 bp in length  
\* 261052 261151: gap of unknown length  
\* 261152 263987: contig of 2836 bp in length  
\* 263988 264087: gap of unknown length  
\* 264088 267533: contig of 3446 bp in length  
\* 267534 267633: gap of unknown length  
\* 267634 279620: contig of 11987 bp in length  
\* 279621 279720: gap of unknown length  
\* 279721 298462: contig of 18742 bp in length  
\* 298463 298562: gap of unknown length  
\* 298563 308821: contig of 10259 bp in length.

FEATURES  
source  
1..308821  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-144P6"  
1..1022  
/note="wgs\_end\_extension  
clone\_end:T7"  
2080..2962  
/note="clone\_boundary  
clone\_end:T7  
site:ECORI  
end\_sequence:RWAXX87TJB"  
2738..4674  
/note="wgs\_contig"

misc\_feature  
misc\_feature  
misc\_feature

Alignment Scores:  
Pred. No.: 7.38e+04 Length: 308821

Score: 88.00 Matches: 33  
Percent Similarity: 42.37% Conservative: 17  
Best Local Similarity: 27.97% Mismatches: 44  
Query Match: 12.43% Indels: 24  
DB: 2 Gaps: 4

US-10-087-573-2 (1-141) x AC118450 (1-308821)  
QY 24 ValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArgAla 43  
Db 39704 GTATATAATACCTCCCTATATAAAGAAAATATATAGAGAAATGCAACTTATTGCTTC 39763  
QY 44 LeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAlaLeuAla 63  
Db 39764 CTACAA-----CAGCAGGGCTCTCCCGGGTATGGCTGCACAGCCAGGCACAG 39811  
QY 64 ThrValThrPro-----LysGlyAlaSerMetLysLeu 74  
Db 39812 CAAGTGCCTCCCTTACAGTCCCAATCCCTAGCTTGATCTGCTCTCTTTAGATTA 39871  
QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
Db 39872 GAACCAAAGGGAGTGAAAGCCTG-----CCAGTCCTTTGGCACCCTTG----- 39913  
QY 95 ArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGlu 114  
Db 39914 -----CGGGTGTCATCTCCACGGTCAAGTATCTCATTATGCGCCAGAAAG 39958  
QY 115 GlyHisProLeuLeuGluLysArgAlaGluTyrPheArgHisLeuArgSerLeu 132  
Db 39959 GGTCATCCAAAGCAGGCAAGAACCACTGGTGGTGGTCACTCAGGCCACTA 40012

Search completed: November 17, 2003, 13:45:56  
Job time : 2504 secs

and is derived by analysis of the total score distribution.

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 17, 2003, 12:04:38 ; Search time 215 Seconds  
(without alignments)  
1770.329 Million cell updates/sec

Title: US-10-087-573-2

Perfect score: 708

Sequence: 1 MESTSTTNFVAENRPTFGE.....RAEYFRHLRLSLKSGVNRLLI 141

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framer\_p2n.model -DEV=xlh  
-O=/cn2\_1/USPTO\_spo1/US10087573/runat\_14112003\_103609\_20663/app\_query.fasta\_1.327  
-DB=N\_Geneseq\_19Jun03 -QFMT=fastap -SUPFI=rrg -WINMATCH=0.1 -LOEFCI=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10087573 @CN 1.1.0 @runat\_14112003\_103609\_20663 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DSV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELXT=7

Database : N\_Geneseq\_19Jun03.\*

1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*  
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	708	100.0	1135	24	ABQ82649	Babesia canis Bcvi
2	680	96.0	1134	24	ABQ82850	Babesia canis Bcvi
C 3	92.5	13.1	2680	24	AB211244	Human polynucleoti
C 4	92.5	13.1	4839	24	ABQ72569	Human MDDT encodin
C 5	92.5	13.1	10123	24	ABK15176	Human REPTR 8 cDNA
C 6	92.5	13.1	10531	25	AA154474	Human cadherin (CA
C 7	92.5	13.1	10759	24	ABA93093	Human dachshous enc
C 8	92	13.0	90	24	ABQ82851	Babesia canis Bcvi
C 9	90	12.7	1386	24	AA563094	Cell death protect
C 10	90	12.7	1657	17	AA730360	Human p57 coding s
C 11	90	12.7	1659	22	AAK54941	Human haematologic
C 12	90	12.7	1659	24	ABK83666	Human cDNA differe
C 13	90	12.7	1659	24	AA563093	Cell death protect
C 14	90	12.7	1842	17	AA730370	GST/truncated huma
C 15	90	12.7	2067	17	AA730369	GST/human p57 fusi
C 16	87	12.3	1007	24	ABQ46392	Oligonucleotide fo
C 17	87	12.3	1007	24	ABQ46393	Oligonucleotide fo
C 18	85.5	12.1	34980	24	ABQ81842	Bifidobacterium lo
C 19	85	12.0	1337	21	AAF13411	Aspergillus oryzae
C 20	83.5	11.8	9213	7	AA60288	Sequence of the HT
C 21	83	11.7	1972	23	ABV21005	Human prostate exp
C 22	83	11.7	1972	23	ABV26849	Human prostate exp
C 23	82.5	11.7	1812	22	AA500717	Chimeric catuq DN
C 24	82.5	11.7	1812	22	AA500718	Chimeric catuq DN
C 25	82	11.6	480	21	AAA82093	N. meningitidis pa
C 26	82	11.6	597	23	AA569040	DNA encoding novel
C 27	82	11.6	10646	22	AA332623	Human genomic DNA
C 28	81.5	11.5	2838	22	AA506738	Polynucleotide seq
C 29	81.5	11.5	2982	24	AA302562	Human kinase poly
C 30	81.5	11.5	3793	24	ABR22334	Prostate cancer-as
C 31	81.5	11.5	9718	17	ABL41480	Attenuated HIV-1 s
C 32	81	11.4	1126	23	ABL28661	Drosophila melanog
C 33	81	11.4	1797	24	AAH77199	Phaeotrichoonis c
C 34	81	11.4	1797	24	AAH77155	Haloperoxidase cod
C 35	81	11.4	3530	24	ABK84630	Human cDNA differe
C 36	81	11.4	3530	24	ABN97376	Gene #3774 used to
C 37	81	11.4	3530	24	ABL69227	Prostate cancer re
C 38	81	11.4	3530	24	ABL69756	Prostate cancer re
C 39	81	11.4	6105	23	AB126682	Drosophila melanog
C 40	80.5	11.4	681	24	ABQ51890	Oligonucleotide fo
C 41	80.5	11.4	681	24	ABQ51891	Oligonucleotide fo
C 42	80.5	11.4	4179	23	ABL13069	Drosophila melanog
C 43	80.5	11.4	6905	23	ABL13068	Drosophila melanog
C 44	80	11.3	1549	24	ABQ33366	Oligonucleotide fo
C 45	80	11.3	1549	24	ABQ33367	Oligonucleotide fo

ALIGNMENTS

RESULT 1  
ABQ82649  
ID ABQ82649 standard; cDNA; 1135 BP.  
XX ABQ82649;  
XX  
XX  
DT 23-DEC-2002 (first entry)  
XX  
DE Babesia canis Bcvi15 15kd protein encoding cDNA SEQ ID NO:1.  
XX  
XX Babesia canis; Bcvi15; 15kd protein; Bcvi32; 32kd protein; infection;  
KW antiparasitic; immunostimulant; vaccine; gene; ss.  
XX  
XX Babesia canis.  
XX  
XX  
FH Key Location/Qualifiers  
FT CDS 75..500



Db 75 ATGGAGTCGACATCAACAGCAACTTTGTCGAGAACCGTCCACCTTTGGTGAG 134  
 QY 21 ThrPheAspValMetArgGluAlaLeuLeuArgValLysSerGluArgLeuAlaMet 40  
 Db 135 ACCTTTCATGATGATGAGGGAAGTTTCTTCGTGTAAGTCTCTGAACGCTTGGCAATG 194  
 QY 41 LeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla 60  
 Db 195 CTCAGAGCGCTTCAGGAATGTGGGTACCGGCTCTCTCTGGCACTGTGTCTCTGGC 254  
 QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgProGln 80  
 Db 255 ATAGCGGCAACGTTAAACCCCAAGGGGCTTCGATGAGCTTAAACACCGCGTCCGAG 314  
 QY 81 SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThr 100  
 Db 315 TCAACGAAGTCTCGGAGCTCAGGAGCTGTCAAGGAATTCGGAATGAATAAGACT 374  
 QY 101 IleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGlu 120  
 Db 375 ATAGTCAGGAATCAGTCGGGTAAACCCCGTTGCCGGAAGCCACCTCTCTTAGAG 434  
 QY 121 LysArgAlaGluThrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140  
 Db 435 AAGCGGCGAGATATTTTCGT-CACCTTAGATCTCTTAAGAGCAAGAGTCAATAGACTC 493  
 QY 141 Ile 141  
 Db 494 ATC 496  
 RESULT 3  
 ABZ11244/c  
 ID ABZ11244 standard; cDNA; 2680 BP.  
 AC ABZ11244;  
 XX  
 XX 20-JAN-2003 (first entry)  
 XX Human polynucleotide SEQ ID NO 126.  
 DE  
 XX Human; genome mapping; gene therapy; food supplement; virus; fungus;  
 KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
 KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
 KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
 KW arthritis; cytostatic; immunomodulator; neotropic; neuroprotective;  
 KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
 KW haemostatic; vulnary; fungicide; antibacterial; virucide; protozoacide;  
 KW antiarthritis; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200270539-A2.  
 XX  
 XX 12-SEP-2002.  
 XX  
 XX 05-MAR-2002; 2002WO-US05095.  
 XX  
 XX 05-MAR-2001; 2001US-0799451.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 XX  
 XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
 PI Wehrman T, Wang J, Wang D, Drmanac RT;  
 XX  
 XX WPI; 2002-759812/82.  
 DR P-PSDB; ABP69027.  
 XX  
 XX New polynucleotides comprising sequences assembled from expressed  
 PT sequence tags (ESTs), useful for treating cell-proliferative,  
 PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or  
 PT platelet or coagulation disorders -

XX  
 PS Claim 1; SEQ ID NO 126; 1012pp + Sequence Listing; English.  
 CC  
 CC The invention relates to an isolated polynucleotide (I) comprising a  
 CC nucleoside sequence selected from any of 948 sequences  
 CC (ABZ119-ABZ1206) or their mature protein coding portion, active domain  
 CC coding protein or complementary sequences. The polynucleotides are useful  
 CC for identifying expressed genes or for physical mapping of human genome.  
 CC The encoded polypeptides (ABP68902-ABP69849) are useful as molecular  
 CC weight markers, as a food supplement, for generating antibodies, in  
 CC medical imaging, screening and diagnostic assays and for treating  
 CC cell-proliferative disorders (cancer), neurodegenerative diseases  
 CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple  
 CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid  
 CC disorders, platelet or coagulation disorders, wound, burns, incision,  
 CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,  
 CC parasitic), arthritis, etc.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcr\_sequences.  
 XX  
 SQ Sequence 2680 BP; 518 A; 863 C; 762 G; 537 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 7.73 Length: 2680  
 Score: 92.50 Matches: 38  
 Percent Similarity: 46.58% Conservative: 30  
 Best Local Similarity: 26.03% Mismatches: 52  
 Query Match: 13.06% Indels: 26  
 DB: 24 Gaps: 6  
 US-10-087-573-2 (1-141) x ABZ11244 (1-2680)  
 QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
 Db 1174 GCAGCACCCACCAACTCCCAAGGAGCTCCACGGCCCTACTAATAGC----- 1124  
 QY 23 AspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
 Db 1123 -----AGGTTGAGTTCAGGTGCCAGGCCCGGCGGTGG 1089  
 QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeu-----ProGlyThrGlyAlaSerAla 60  
 Db 1087 GTGATATACCGCTCACAGGCACTGTGGCACTCCGGGAACCCAGGCGAGCCCGTGTCT 1028  
 QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgProGln 80  
 Db 1027 ATCAGCTTCAGCTCAGCTCCGCTGCTTCCCGCGGTCCCGCCCGCCCGCCCGGAG 968  
 QY 81 ---SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArg----- 95  
 Db 967 GTGGCTGTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 908  
 QY 96 -----GluMetAsnLysThrIleSerGlnGluSerAlaArgVal-AsnHisArg 111  
 Db 907 TGGTTAATACCAAAATAGGGGAAGAGGTGGCAGAGAA---TACAGAACCCAGGCGCATCG 851  
 QY 111 GLeuProGluGlyHis-----ProLeuLeuGluLysArgAlaGluThrPheArgHisLe 129  
 Db 850 GCACCCCATCTCTCATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 791  
 QY 129 uArgSerLeuLysSer 134  
 Db 790 TCGGCACTTGGAAAGT 775  
 RESULT 4  
 ABQ72569/c  
 ID ABQ72569 standard; cDNA; 4639 BP.  
 XX  
 XX AC ABQ72569;  
 XX DT 03-SEP-2002 (first entry)  
 XX



DE Human MDDT encoding cDNA SEQ ID NO 121.  
XX  
KW Human; MDDT; disease detection and treatment molecule polynucleotide;  
KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;  
KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;  
KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;  
KW hepatotropic; antiinflammatory; antipsoriatic; cytostatic; anti-HIV;  
KW antiatherogenic; antianemic; antiasthmatic; antiarteriosclerotic; antigout;  
KW neuroprotective; antirheumatic; antiarthritic; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200240715-A2.  
XX  
PD 23-MAY-2002.  
XX  
PF 06-SEP-2001; 2001WO-US27628.  
XX  
PR 06-SEP-2000; 2000US-230505P.  
PR 06-SEP-2000; 2000US-230514P.  
PR 06-SEP-2000; 2000US-230515P.  
PR 06-SEP-2000; 2000US-230517P.  
PR 06-SEP-2000; 2000US-230518P.  
PR 06-SEP-2000; 2000US-230519P.  
PR 06-SEP-2000; 2000US-230595P.  
PR 06-SEP-2000; 2000US-230597P.  
PR 06-SEP-2000; 2000US-230598P.  
PR 06-SEP-2000; 2000US-230599P.  
PR 06-SEP-2000; 2000US-230610P.  
PR 06-SEP-2000; 2000US-230865P.  
PR 06-SEP-2000; 2000US-230988P.  
PR 06-SEP-2000; 2000US-230989P.  
PR 07-SEP-2000; 2000US-230951P.  
PR 07-SEP-2000; 2000US-231163P.  
PR 07-SEP-2000; 2000US-231167P.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;  
PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;  
PI Momiya MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;  
PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;  
PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;  
XX  
WPI: 2002-527544/56.  
DR P-PSDB; ABP51351.  
XX  
PT Novel human disease detection and treatment polypeptide, useful in  
PT diagnosis, prevention or treatment of cell proliferative disorders e.g.  
PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder  
PT e.g. AIDS -  
XX  
PS Claim 1; Page 375-376; 618pp; English.

XX The invention relates to an isolated human disease detection and  
XX treatment (MDDT) polypeptide (I) selected from a polypeptide having a  
XX sequence selected from 254 sequences (ABP51231-ABP51484) given in the  
XX specification, a naturally occurring polypeptide comprising a sequence  
XX having at least 90% identity to (I) or a biologically active or  
XX immunogenic fragment of (I). (I) is useful for screening a compound for  
XX effectiveness as an agonist or antagonist, for screening a compound that  
XX specifically binds (I) or modulates the activity of (I), and for  
XX preparing a polyclonal or monoclonal antibody by hybridoma technology.  
XX Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for  
XX screening a compound for effectiveness in altering expression of a target  
XX polynucleotide comprising. Oligonucleotides and antibodies are useful for  
XX detecting MDDT in a sample or for assessing toxicity of a test compound,  
XX in a diagnostic test for a condition or a disease associated with the  
XX expression of MDDT in a biological sample, for detecting (I) in a sample,  
XX and for purifying (I) from a sample. A composition comprising (I), an  
XX agonist or antagonist is useful for treating a disease or condition  
XX associated with decreased or increased expression of functional MDDT.  
XX (I) or (II) are useful for diagnosing, treating or preventing disorders

CC associated with aberrant expression of MDDT, where the disorders are  
CC selected from a cell proliferative disorder such as arteriosclerosis,  
CC cirrhosis, hepatitis, psoriasis, and cancer and an  
CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,  
CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or  
CC rheumatoid arthritis. (II) are useful for creating knockin humanised  
CC animals or transgenic animals to model human diseases, in somatic or  
CC germline gene therapy, to generate a transcript image of a tissue or cell  
CC type, for detecting differences in the chromosomal location due to  
CC translocation or inversion among normal, carrier or affected individuals  
CC and as hybridisation probes for mapping naturally occurring genomic  
CC sequences.  
XX  
SQ Sequence 4839 BP; 946 A; 1467 C; 1418 G; 1008 T; 0 other;

Alignment Scores:  
Pred. No.: 16.8 Length: 4839  
Score: 92.50 Matches: 38  
Percent Similarity: 46.58% Conservative: 30  
Best Local Similarity: 26.03% Mismatches: 52  
Query Match: 13.06% Indels: 26  
DB: 24 Gaps: 6  
US-10-087-573-2 (1-141) x ABQ72569 (1-4839)  
QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
Db 2839 GCAGCACACACACAACTCCCAAGAGGCTGCCAGGCCCTACTAATAGC----- 2789  
QY 23 AspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
Db 2788 -----AGGTTGAGGTCAGTGCCAGGCCCGCCGCTGCT 2753  
QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeu-----ProGlyThrGlyAlaSerAla 60  
Db 2752 GTGATATCCAGGTCACAGGCACTGTGGCACTCCGGGAACCCAGGAGGCCCGCTCT 2693  
QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgProGln 80  
Db 2692 ATCACTCCAGCTCAGCTCCGCTGCTTCCCGCGGCTCCCGCCCGCCCGCCAGAG 2633  
QY 81 ---SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArg----- 95  
Db 2632 GTGGCTGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2573  
QY 96 -----GluMetAsnLysThrLysSerGlnGluSerAlaArgVal-AsnHisArg 111  
Db 2572 TGGTTAATACCAATAGGGGAAGAGGTGGCAAGGAA---TACAGAACCCAGCCATCG 2516  
QY 111 GlnProGluGlyHis-----ProLeuLeuGluLysArgAlaGluTyrPheArgHisLe 129  
Db 2515 GCACCCCATCTCATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2456  
QY 129 uArgSerLeuLysSer 134  
Db 2455 TCGGCACTTGGAGT 2440  
RESULT 5  
ABK15176/c  
ID ABK15176 standard; DNA; 10123 BP.  
XX  
AC ABK15176;  
XX  
DT 23-APR-2002 (first entry)  
XX  
DE Human REPTR 8 cDNA sequence.  
KW REPTR; human; antiinflammatory; cytostatic; immunosuppressive;  
KW antiviral; anti-HIV; antiarthritic; anticonvulsant; nootropic;  
KW neuroprotective; antiatherogenic; antibody; immunogen; endometriosis;  
KW gastrointestinal disorder; gastritis; oesophageal carcinoma;  
KW Crohn's disease; irritable bowel syndrome; ulcerative colitis;  
KW endocrine disorder; hypothalamus disorder; Kallman's disease;



XX The invention relates to a novel method for identifying a candidate p53  
 CC pathway modulating agent. The method comprises providing an assay system  
 CC comprising a purified cadherin (CAD) polypeptide or nucleic acid, or  
 CC their functionally active fragment or derivative. The method is useful  
 CC for identifying a candidate p53 pathway modulating agent, modulating a  
 CC p53 pathway of a cell and for diagnosing a disease in a patient. In  
 CC particular, the disease is cancer, e.g. breast cancer, colon cancer,  
 CC kidney cancer, lung cancer or cancer of the ovary, which has an  
 CC expression level of greater than 25%. The identified modulators are  
 CC useful as targets for novel therapeutics. This polynucleotide sequence  
 CC represents a cadherin (CAD) nucleic acid of the invention.  
 XX  
 XX Sequence 10531 BP; 1980 A; 3340 C; 3141 G; 2070 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 46.8 Length: 10531  
 Score: 92.50 Matches: 38  
 Percent Similarity: 46.58% Conservative: 30  
 Best Local Similarity: 26.03% Mismatches: 52  
 Query Match: 13.06% Indels: 26  
 DB: 25 Gaps: 6  
 US-10-087-573-2 (1-141) x AAL54474 (1-10531)  
 QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
 DB 9037 GCAAGCACCACCACTCCCAAGGAGCTGCCAGGCCCTACTAATAGC----- 8987  
 QY 23 AspValMetArgGluAlaLeuLeuValLysSerSerGluArgLeuAlaMetLeuArg 42  
 DB 8986 -----AGGTTAGGTTCAGGTGCCAGGCCCAAGGCGGTGGTGG 8951  
 QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeu-----ProGlyThrGlyAlaSerAla 60  
 DB 8950 GTGATATCCACGTCACAGGCACTGGGCATCCGGAAACAGGAGGCGCCCGTCT 8891  
 QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgProGln 80  
 DB 8890 ATCACCCTCCAGCTCAGCTCCCGTGTGCTTCCCGCGGTCGCGGCCCAAGGAG 8831  
 QY 81 ---SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArg----- 95  
 DB 8830 GTGGCTGTTCCGCTGCTGCCGACGTGTCACCCGAGGTACAGGCTCCTGTAGTC 8771  
 QY 96 -----GluMetAsnLysThrIleSerGlnGluSerAlaArgVal-AsnHisArg 111  
 DB 8770 TGGTTAATACCAAAATAGGGGAAGAGGTGGCAAGGAA---TACAGAACCAAGGCAATCG 8714  
 QY 111 gLeuProGluGlyHis-----ProLeuLeuGluLysArgAlaGluTyrPheArgHisLe 129  
 DB 8713 GCACCCCAATCTCTATCTGTGGCTGCACGTGACCCCAAGCTGTGGCCAGCGCGGCACCT 8654  
 QY 129 uArgSerLeuLysSer 134  
 DB 8653 TCGGGCACTTGAAGT 8638  
 RESULT 7  
 ID ABA93093/c  
 ID ABA93093 standard; cDNA; 10759 BP.  
 XX  
 AC ABA93093;  
 XX  
 DT 12-APR-2002 (first entry)  
 XX  
 DE Human dachshous encoding cDNA SEQ ID NO:1.  
 XX  
 KW Dachshous; Freeman-Sheldon syndrome; congenital anamorphosis syndrome;  
 XX human; gene; ss.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers

CDS 411..10307  
 FT /\*tag= a  
 FT /product= "dachshous protein"  
 XX  
 PN JP2001327295-A.  
 XX  
 PD 27-NOV-2001.  
 XX  
 PF 27-JUL-2000; 2000JP-0226291.  
 XX  
 PR 14-MAR-2000; 2000JP-0069599.  
 XX  
 PA (KAZU-) 2H KAZUSA DNA KENKYUSHO.  
 XX  
 DR WI; 2002-127071/17.  
 DR P-PSDB; ABB05430.  
 XX  
 PT A human dachshous gene and a protein encoded by the gene -  
 XX  
 PS Claim 2; Page 10-27; 40pp; Japanese.  
 XX  
 CC The present sequence encodes a human dachshous protein. The dachshous  
 CC gene can be used for diagnosing or treating Freeman-Sheldon syndrome  
 CC or congenital anamorphosis syndrome.  
 XX  
 SQ Sequence 10759 BP; 2008 A; 3412 C; 3256 G; 2083 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 48.1 Length: 10759  
 Score: 92.50 Matches: 38  
 Percent Similarity: 46.58% Conservative: 30  
 Best Local Similarity: 26.03% Mismatches: 52  
 Query Match: 13.06% Indels: 26  
 DB: 24 Gaps: 6  
 US-10-087-573-2 (1-141) x ABA93093 (1-10759)  
 QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
 DB 9265 GCAAGCACCACCACTCCCAAGGAGCTGCCAGGCCCTACTAATAGC----- 9215  
 QY 23 AspValMetArgGluAlaLeuLeuValLysSerSerGluArgLeuAlaMetLeuArg 42  
 DB 9214 -----AGGTTAGGTTCAGGTGCCAGGCCCAAGGCGGTGGTGG 9179  
 QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeu-----ProGlyThrGlyAlaSerAla 60  
 DB 9178 GTGATATCCACGTCACAGGCACTGGGCATCCGGAAACAGGAGGCGCCCGTCT 9119  
 QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgProGln 80  
 DB 9118 ATCACCCTCCAGCTCAGCTCCCGTGTGCTTCCCGCGGTCGCGGCCCAAGGAG 9059  
 QY 81 ---SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArg----- 95  
 DB 9058 GTGGCTGTTCCGCTGCTGCCGACGTGTCACCCGAGGTACAGGCTCCTGTAGTC 8999  
 QY 96 -----GluMetAsnLysThrIleSerGlnGluSerAlaArgVal-AsnHisArg 111  
 DB 8998 TGGTTAATACCAAAATAGGGGAAGAGGTGGCAAGGAA---TACAGAACCAAGGCAATCG 8942  
 QY 111 gLeuProGluGlyHis-----ProLeuLeuGluLysArgAlaGluTyrPheArgHisLe 129  
 DB 8941 GCACCCCAATCTCTATCTGTGGCTGCACGTGACCCCAAGCTGTGGCCAGCGCGGCACCT 8882  
 QY 129 uArgSerLeuLysSer 134  
 DB 8881 TCGGGCACTTGAAGT 8866  
 RESULT 8  
 ID ABA92651  
 ID ABA92651 standard; DNA; 90 BP.  
 XX

AC ABQ82651;  
 CC  
 DT 23-DEC-2002 (first entry)  
 XX  
 DE Babesia canis Bcvir oligonucleotide PCR primer SEQ ID NO:5.  
 XX  
 KW Babesia canis; Bcvir15; 15kd protein; Bcvir32; 32kd protein; infection;  
 KW antiparasitic; immunostimulant; vaccine; PCR primer; ss.  
 XX  
 XX Babesia canis.  
 OS  
 XX  
 XX EP1238983-A1.  
 PN  
 XX  
 PD 11-SEP-2002.  
 XX  
 XX 04-MAR-2002; 2002EP-0075830.  
 PF  
 XX 06-MAR-2001; 2001EP-0200816.  
 PR  
 XX (ALKU ) AKZO NOBEL NV.  
 PA  
 XX Schettlers TPM, Carcy BPD, Drakulovski PR, Gorenflot AF;  
 PI  
 XX WPI; 2002-724917/79.  
 DR  
 XX  
 XX Novel Babesia canis associated protein and nucleic acid encoding the  
 PT protein, useful in a vaccine and in the manufacture of vaccines for  
 PT combating Babesia canis infections. -  
 XX  
 XX Example 2; Page 12; 41pp; English.  
 PS  
 XX  
 XX The present invention describes a Babesia canis associated protein (I),  
 CC comprising a Bcvir15 protein of 15 kD molecular weight (MW) and having  
 CC a sequence of at least 80% homology to a sequence (ABP53714) of 141  
 CC amino acids, or a Bcvir32 protein of 32 kD MW and having a sequence of  
 CC at least 80% homology to a sequence (ABP53715) of 285 amino acids, or  
 CC their immunogenic fragments. (I) have antiparasitic and immunostimulant  
 CC activities, and can be used in vaccines. (I) can also be used for the  
 CC preparation of a vaccine for combating B. canis infections. (I) is also  
 CC useful in a diagnostic test for the detection of antibodies against  
 CC B. canis associated antigenic material. The present sequence represents  
 CC a PCR primer for Babesia canis Bcvir which is used in an example from  
 CC the present invention.  
 XX  
 SQ Sequence 90 BP; 26 A; 27 C; 19 G; 18 T; 0 other;

KW nutritional condition; peripheral nervous system disorder; ischaemia;  
 KW diabetic neuropathy; autoimmune haemolytic anaemia; respiratory system;  
 KW oral cavity; gastrointestinal tract; liver; cirrhosis; pancreatitis;  
 KW polycystic renal disease; urinary tract; genitalia; endometriosis;  
 KW breast; chronic mastitis; thyroid gland; Hashimoto's thyroiditis;  
 KW adrenal gland; skin; psoriasis; muscular atrophy; bone marrow;  
 KW osteoporosis; cancer; autoimmune disease; ss.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX WO200176532-A2.  
 PN  
 XX  
 PD 18-OCT-2001.  
 XX  
 XX 09-APR-2001; 2001WO-US11655.  
 PF  
 XX 11-APR-2000; 2000US-0547596.  
 PR  
 XX (COGE-) COGENT NEUROSCIENCE INC.  
 PA  
 XX Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC, Lo DC;  
 PI  
 XX WPI; 2002-017408/02.  
 DR  
 XX P-PSDB; AAU69691.  
 DR  
 XX  
 XX Novel nucleic acids referred as protective sequences and their encoded  
 PT products for diagnosing, treating diseases involving cell death,  
 PT including neurological disorders e.g. stroke and for identifying  
 PT modulators -  
 XX  
 XX Claim 2; Figure 10A; 256pp; English.  
 PS  
 XX The invention relates to isolated protective sequence polypeptides (I)  
 CC and polynucleotides (II). (I) is useful for transferring a protective  
 CC sequence into a cell, which delays and/or prevents the cell from  
 CC undergoing cell death. Protective sequences, their products or  
 CC antibodies are useful diagnostically, prophylactically, therapeutically  
 CC or as targets for treatment and diagnosis of conditions, disorders or  
 CC diseases involving cell death. The protective sequences and their  
 CC products are useful for preventing or treating disorders of the central  
 CC nervous system including neurological and psychiatric conditions,  
 CC cerebral oedema, infections such as meningitis, degenerative diseases  
 CC such as Alzheimer's and motor neuron disease, demyelinating diseases such  
 CC as multiple sclerosis, nutritional conditions, disorders of the  
 CC peripheral nervous system including diabetic neuropathy, disorders  
 CC which cause cell death in organ systems including blood vessels, heart  
 CC (ischaemia), blood cells (autoimmune haemolytic anaemia), respiratory  
 CC system (asthma), oral cavity, gastrointestinal tract, liver (cirrhosis),  
 CC pancreatitis, polycystic renal disease, urinary tract, genitalia  
 CC (congenital anomalies), endometriosis, breast (chronic mastitis), thyroid  
 CC gland (Hashimoto's thyroiditis), adrenal gland, skin (psoriasis),  
 CC musculoskeletal system (muscular atrophy), bone marrow or bone  
 CC (osteoporosis). The compositions promote cell death and are useful for  
 CC treating and/or ameliorating cancer and autoimmune diseases. The  
 CC compounds are further useful for treating physiological impacts on  
 CC organs caused by infection which induce cell death. (I) is useful to  
 CC raise an immune response, as a reagent in assays designed to  
 CC quantitatively determine levels of the protein in biological fluids, as  
 CC markers for tissues in which the corresponding protein is expressed and  
 CC to isolate receptors or ligands. AAS62967-AAS63142 represent the  
 CC protective polynucleotide sequences as described in the invention.  
 XX  
 SQ Sequence 1386 BP; 268 A; 416 C; 460 G; 242 T; 0 other;

Alignment Scores:  
 Pred. No.: 6 Length: 1386  
 Score: 90.00 Matches: 40  
 Percent Similarity: 41.54% Conservative: 14  
 Best Local Similarity: 30.77% Mismatches: 41  
 Query Match: 12.71% Indels: 35  
 DB: 24 Gaps: 6

US-10-087-573-2 (1-141) x AAS63094 (1-1386)

QY 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPhe 18  
 DB 35 ATGGAGTCGACATCAACACACGACCAACTTTGTTGCCGAGACCGCCACCTTT 88

RESULT 9  
 AAS63094/C  
 ID AAS63094 standard; cDNA; 1386 BP.  
 XX  
 XX AAS63094;  
 AC  
 XX  
 XX 29-JAN-2002 (first entry)  
 DT  
 XX  
 XX Cell death protective sequence CNI-00729, ORF #1.  
 DE  
 XX Human; protective sequence; cell death; cerebral oedema; infection;  
 KW meningitis; degenerative disease; Alzheimer's disease; heart disease;  
 KW motor neuron disease; demyelinating disease; multiple sclerosis; asthma;

Qy 8 ThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg--- 26  
Dy 401 ACACGCTGGTGTGGCCCTCCAGGCTGACAGGGCTCCCGAGGGGCGACATCAGGCC 342  
Qy 27 -----GluAlaLeuLeuArgValLysSerGluArgLeuAlaMetLeuArg 42  
Dy 341 CCATCTGGGANTCCACACCATGACTGTGCAGTCTCCGAGGCCACTGGCAATGACGTTG 282  
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
Dy 281 TCATTG-----TGGGGCACCAGGCGATGTCTAGCACAGGGGCTGTGTGCCACAG 231  
Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
Dy 230 ACCGTGGGGCGCATCTTCTGCCACAGTCCAGTCTTGGCCAGGGGCGACACAGGAAGGCC 171  
Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysile 94  
Dy 170 CTTCCCGCG-----CTGGCCTCACAGATC 147  
Qy 95 ArgGluMetAsn-----LysThrIleSerGlnGlu-SerAlaArgVa 108  
Dy 146 AGGGCCACAACTTAGGGTTGACAGCACAGAAGCCATGTCCAGGTGTCTGTGAGACG 87  
Qy 108 lAsnHisArgLeuProGluGlyHisPro 117  
Dy 86 CGCACATCTTCATAGCACTGGTCGGCCT 59  
RESULT 10  
AAT30360/c  
ID AAT30360 standard; cDNA; 1657 BP.  
XX AC AAT30360;  
XX DT 26-NOV-1996 (first entry)  
XX DE Human p57 coding sequence.  
XX KW WD40 repeated structure region; human p57 protein; dimer;  
XX KW anti-cancer; anti-HIV agent; actin; cell movement; ds.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
XX CDS 100..1485  
XX FT /\*tag= a  
XX FT /product= p57  
XX PN JP08119996-A.  
XX PD 14-MAY-1996.  
XX PF 21-OCT-1994; 94JP-0282743.  
XX PR 21-OCT-1994; 94JP-0282743.  
XX PA (NIBS ) JAPAN TOBACCO INC.  
XX WP; 1996-283507/29.  
XX DR P-PSDB; AAR98341.  
XX PT A new protein, p57, comprising WD40 repeat region - used for  
XX development of anti-cancer and anti-HIV agents  
XX Claim 9; Page 23-25; 51pp; Japanese.  
XX CC This sequence encodes the human p57 protein. p57 contains a leucine  
XX rich C-terminal peptide which comprises a leucine residue after  
XX each 7 amino acids (see also R983340) and a WD40 repeated structure  
XX region (see also AAR98334-38) containing five WD40 regions. p57 forms a  
XX dimer. It can be used in the development of an anti-cancer agent and  
XX an anti-HIV agent. It can also combine with actin in the control of

CC cell movement.  
XX SQ Sequence 1657 BP; 359 A; 497 C; 511 G; 290 T; 0 other;  
Alignment Scores: 7.58 Length: 1657  
Pred. No.: 90.00 Matches: 40  
Score: 41.54% Conservative: 14  
Best Local Similarity: 30.77% Mismatches: 41  
Query Match: 12.71% Indels: 35  
DB: 17 Gaps: 6  
US-10-087-573-2 (1-141) x AAT30360 (1-1657)  
Qy 8 ThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg--- 26  
Dy 500 ACACGCTGGTGTGGCCCTCCAGGCTGACAGGGCTCCCGAGGGGCGACATCAGGCC 441  
Qy 27 -----GluAlaLeuLeuArgValLysSerGluArgLeuAlaMetLeuArg 42  
Dy 440 CCATCCGGGATCTCCACACCATGATGTGCAGTCTCCGAGGCCACTGGCAATGACGTTG 381  
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
Dy 380 TCATTG-----TGGGGCACCAGGCGATGTCTAGCACAGGGGCTGTGTGCCACAG 330  
Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
Dy 329 ACCGTGGGGCGCATCTTCTGCCACAGTCCAGTCTTGGCCAGGGGCGACACAGGAAGGCC 270  
Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysile 94  
Dy 269 CTTCCCGCG-----CTGGCCTCACAGATC 246  
Qy 95 ArgGluMetAsn-----LysThrIleSerGlnGlu-SerAlaArgVa 108  
Dy 245 AGGGCCACAACTTAGGGTTGACAGCACAGAAGCCATGTCCAGGTGTCTGTGAGACG 186  
Qy 108 lAsnHisArgLeuProGluGlyHisPro 117  
Dy 185 CGCACATCTTCATAGCACTGGTCGGCCT 158  
RESULT 11  
AAR54941/c  
ID AAR54941 standard; cDNA; 1659 BP.  
XX AC AAR54941;  
XX DT 13-NOV-2001 (first entry)  
XX DE Human haematological malignancy-related antigen coding sequence #666.  
XX KW Human; cytostatic; vascular; gene therapy; vaccine; lymphoma;  
XX KW haematological malignancy; antigen; chronic lymphocytic leukaemia;  
XX KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma; ss.  
XX OS Homo sapiens.  
XX PN WO200164896-A2.  
XX PD 07-SEP-2001.  
XX PF 01-MAR-2001; 2001WO-US07272.  
XX PR 01-MAR-2000; 2000US-0186126.  
XX PR 17-MAR-2000; 2000US-0190479.  
XX PR 27-APR-2000; 2000US-0200545.  
XX PR 28-APR-2000; 2000US-0200303.  
XX PR 28-APR-2000; 2000US-0200779.  
XX PR 01-MAY-2000; 2000US-0200999.  
XX PR 04-MAY-2000; 2000US-0202084.  
XX PR 22-MAY-2000; 2000US-0206201.  
XX PR 14-JUL-2000; 2000US-0218950.

PR 03-AUG-2000; 2000US-0222903.  
PR 04-AUG-2000; 2000US-0223416.  
PR 07-AUG-2000; 2000US-0223378.  
XX (CORI-) CORIXA CORP.  
XX Gaiger A, Algate PA, Mannion J;  
XX WPI; 2001-514842/56.  
XX Compositions and methods for the detection of hematological  
PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular  
PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -  
XX  
PS Claim 31; Pages 504-505; 1252pp; English.  
XX  
XX The present invention relates to compositions and methods for the  
CC detection, diagnosis and therapy of haematological malignancies. The  
CC present sequence is the coding sequence of a human haematological  
CC malignancy related antigen. The methods of the present invention comprise  
CC detecting the presence of haematological malignancy related antigen(s) in  
CC a sample obtained from the patient (an increased level of the  
CC polypeptide, compared to an unaffected individual, is indicative of an  
CC increased risk). Haematological malignancies which can be treated using  
CC the present invention are chronic lymphocytic leukaemia, lymphoma,  
CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B  
CC cell non-Hodgkin's lymphoma.  
XX  
SQ Sequence 1659 BP; 360 A; 497 C; 511 G; 291 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 7.6 Length: 1659  
Score: 90.00 Matches: 40  
Percent Similarity: 41.54% Conservative: 14  
Best Local Similarity: 30.77% Mismatches: 41  
Query Match: 12.71% Indels: 35  
DB: 22 Gaps: 6  
  
US-10-087-573-2 (1-141) x AAK54941 (1-1659)  
Qy 8 ThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg --- 26  
Db 501 ACACGCTGGTGCGCCCTCCAGGCGTACGACGCGCTCCCGAGGGCGACATCAGGCC 442  
Qy 27 -----GluAlaLeuLeuArgValIysSerSerGluArgLeuAlaMetLeuArg 42  
Db 441 CCATCCGGGATCTCCACACCATGACTGTGCGAGTCTCGGAGCCACTGGCAATGACGCTG 382  
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla ----- 58  
Db 381 TCATTTG -----TGCGGCACACGCGGATGCTAGCACAGGGGCTGTGTGCCACAG 331  
Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
Db 330 ACCGTGGGCGCATCTTGTCCACACGCTCCAGTCTTCCAGGGCGACACAGGAGGCC 271  
Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
Db 270 CCTCCCGCG -----CTGGCCTCACAGATC 247  
Qy 95 ArgGluMetAsn -----LysThrIleSerGlnGlu-SerAlaArgVa 108  
Db 246 AGGCCACAACTTAGGTTGACAGCAGAGGACCACTGTCCAGGGTGCTGTGAGACG 187  
Qy 108 LAsnHisArgLeuProGluGlyHisPro 117  
Db 186 CGCACATCTTCATAGCACTGGTCGGCCT 159  
  
RESULT 12  
ABK83666/c  
ID ABK83666 standard; cDNA; 1659 BP.  
XX  
AC ABK83666;

XX  
DT 14-AUG-2002 (first entry)  
XX Human cDNA differentially expressed in granulocytic cells #237.  
DE  
XX  
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;  
KW viral infection; parasitic infection; protozoal infection;  
KW fungal infection; sterile inflammatory disease; psoriasis;  
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
KW adult respiratory distress syndrome; inflammatory bowel disease;  
KW Crohn's disease; ulcerative colitis; periodontal disease;  
KW granulocyte activation; chronic inflammation; allergy.  
XX  
OS Homo sapiens.  
XX  
PN WO200228999-A2.  
PD 11-APR-2002.  
XX  
PF 03-OCT-2001; 2001WO-US30821.  
PR 03-OCT-2000; 2000US-237189P.  
XX (GENE-) GENE LOGIC INC.  
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
XX  
XX WPI; 2002-435328/46.  
DR  
XX  
XX Detecting granulocyte activation by detecting differential expression  
PT of genes associated with granulocyte activation, which serves as  
PT diagnostic markers that is useful for monitoring disease states and  
XX drug toxicity -  
PS Claim 1; SEQ ID No 237; 114pp; English.  
XX  
XX The invention relates to detecting (M1) granulocyte (GC) activation  
CC (GCA), by detecting the level of expression of gene(s) identified by  
CC DNA chip analysis as given in the specification, and comparing  
CC the expression level to an expression level in an unactivated  
CC GC, where differential expression of Gs is indicative of GCA.  
CC Also included are modulating (M2) GA by contacting GC with an agent  
CC that alters the expression of at least one gene in Gs; (2) screening (M3)  
CC for an agent capable of modulating GCA or an inflammation (especially  
CC chronic) in a tissue, an allergic response in a subject, exposure of a  
CC subject to a pathogen or sterile inflammatory disease using the  
CC gene expression profile; (3) detecting (M4) an inflammation (especially  
CC chronic) in a tissue, an allergic response in a subject, exposure of a  
CC subject to a pathogen or sterile inflammatory disease, by detecting the  
CC level of expression in a sample of the tissue of gene(s) from Gs, where  
CC the level of expression of the gene is indicative of inflammation;  
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,  
CC an allergic response in a subject, exposure of a subject to a pathogen  
CC or sterile inflammatory disease, by contacting a tissue having  
CC inflammation with an agent that modulates the expression of gene(s)  
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for  
CC modulating GA; M3 is useful for screening an agent capable of modulating  
CC GCA preferably in an inflammation in a tissue; M4 is useful for  
CC detecting an inflammation (especially chronic) in a tissue, an allergic  
CC response in a subject, exposure of a subject to a pathogen or sterile  
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis, renal  
CC glomerulonephritis, ARDS, thrombosis, cardiac reperfusion injury, renal  
CC reperfusion injury, asthma, adult respiratory distress syndrome,  
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,  
CC periodontal disease; also bacterial infection, viral infection,  
CC parasitic infection, protozoal infection, fungal infection and M5 is  
CC useful for treating one of the above conditions. The present  
CC sequence represents a gene differentially expressed in granulocytes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1659 BP; 360 A; 497 C; 511 G; 291 T; 0 other;  
Alignment Scores:  
Pred. No.: 7.6 Length: 1659  
Score: 90.00 Matches: 40  
Percent Similarity: 41.54% Conservative: 14  
Best Local Similarity: 30.77% Mismatches: 41  
Query Match: 12.71% Indels: 35  
DB: 24 Gaps: 6  
US-10-087-573-2 (1-141) x ABR83666 (1-1659)  
QY 8 ThrAsnPhEValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg--- 26  
DB 501 ACACGCTTGGTGGCTCCAGGTGACGAGGCTCCCGAGGGGAGCATCAGGCC 442  
QY 27 -----GluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
DB 441 CCATCGGGATCTCCACACCATGACTGTGCAGTCTCGAGCCACTGGCAATGACGTTG 382  
QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
DB 381 TCATTG-----TGGCGGCACCGAGGTCTAGCACAGGGGCTGTGGCCACAG 331  
QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
DB 330 ACGTGGGGCATCTTGTCCACAGTCCAGTCTCCAGGGGAGGAGGCC 271  
QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysLeu 94  
DB 270 CTTCCCGCC-----CTGGCTTCACAGATC 247  
QY 95 ArgGluMetAsn-----LysThrIleSerGlnGlu-SerAlaArgVa 108  
DB 246 AGGCCACAACTAGGTTGACAGCACAGAGCCACTGCCAGGTGCTGTGAGAGC 187  
QY 108 lAsnHisArgLeuProGluGlyHisPro 117  
DB 186 CGACATCTTCATAGCACTGGTGGGCT 159  
RESULT 13  
AAS63093/C  
ID AAS63093 standard; cDNA; 1659 BP.  
XX AC AAS63093;  
XX DT 29-JAN-2002 (first entry)  
XX DE Cell death protective sequence CNI-00729.  
XX KW Human; protective sequence; cell death; cerebral oedema; infection;  
KW meningitis; degenerative disease; Alzheimer's disease; heart disease;  
KW motor neuron disease; demyelinating disease; multiple sclerosis; asthma;  
KW nutritional condition; peripheral nervous system disorder; ischaemia;  
KW diabetic neuropathy; autoimmune haemolytic anaemia; respiratory system;  
KW oral cavity; gastrointestinal tract; liver; cirrhosis; pancreatitis;  
KW polycystic renal disease; urinary tract; genitalia; endometriosis;  
KW breast; chronic mastitis; thyroid gland; Hashimoto's thyroiditis;  
KW adrenal gland; skin; psoriasis; muscular atrophy; bone marrow;  
KW osteoporosis; cancer; autoimmune disease; ss.  
XX OS Homo sapiens.  
XX PN WO200176532-A2.  
XX PD 18-OCT-2001.  
XX PF 09-APR-2001; 2001WO-US11655.  
XX PR 11-APR-2000; 2000US-0547596.  
XX PA (COGE-) COGENT NEUROSCIENCE INC.

XX PI Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC, Lo DC;  
XX WPI; 2002-017408/02.  
XX Novel nucleic acids referred as protective sequences and their encoded  
PT products for diagnosing, treating diseases involving cell death,  
PT including neurological disorders e.g. stroke and for identifying  
XX modulators -  
XX Claim 2; Figure 1G; 256pp; English.  
XX The invention relates to isolated protective sequence polypeptides (I)  
CC and polynucleotides (II) (II) is useful for transferring a protective  
CC sequence into a cell, which delays and/or prevents the cell from  
CC undergoing cell death. Protective sequences, their products or  
CC antibodies are useful diagnostically, prophylactically, therapeutically  
CC or as targets for treatment and diagnosis of conditions, disorders or  
CC diseases involving cell death. The protective sequences and their  
CC products are useful for preventing or treating disorders of the central  
CC nervous system including neurological and psychiatric conditions,  
CC cerebral oedema, infections such as meningitis, degenerative diseases  
CC such as Alzheimer's and motor neuron disease, demyelinating diseases such  
CC as multiple sclerosis, nutritional conditions, disorders of the  
CC peripheral nervous system including diabetic neuropathy, disorders  
CC which cause cell death in organ systems including blood vessels, heart  
CC (ischaemia), blood cells (autoimmune haemolytic anaemia), respiratory  
CC system (asthma), oral cavity, gastrointestinal tract, liver (cirrhosis),  
CC pancreatitis, polycystic renal disease, urinary tract, genitalia  
CC (congenital anomalies), endometriosis, breast (chronic mastitis), thyroid  
CC gland (Hashimoto's thyroiditis), adrenal gland, skin (psoriasis),  
CC musculoskeletal system (muscular atrophy), bone marrow or bone  
CC (osteoporosis). The compositions promote cell death and are useful for  
CC treating and/or ameliorating cancer and autoimmune diseases. The  
CC compounds are further useful for treating physiological impacts on  
CC organs caused by infection which induce cell death. (I) is useful to  
CC raise an immune response, as a reagent in assays designed to  
CC quantitatively determine levels of the protein in biological fluids, as  
CC markers for tissues in which the corresponding protein is expressed and  
CC to isolate receptors or ligands. AAS62967-AAS63142 represent the  
XX protective polynucleotide sequences as described in the invention.  
SQ Sequence 1659 BP; 331 A; 518 C; 514 G; 296 T; 0 other;

Alignment Scores:  
Pred. No.: 7.6 Length: 1659  
Score: 90.00 Matches: 40  
Percent Similarity: 41.54% Conservative: 14  
Best Local Similarity: 30.77% Mismatches: 41  
Query Match: 12.71% Indels: 35  
DB: 24 Gaps: 6  
US-10-087-573-2 (1-141) x AAS63093 (1-1659)  
QY 8 ThrAsnPhEValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg--- 26  
DB 540 ACACGCTTGGTGGCTCCAGGTGACGAGGCTCCCGAGGGGAGCATCAGGCC 481  
QY 27 -----GluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
DB 480 CCATCGGGATCTCCACACCATGACTGTGCAGTCTCGAGGCCACTGGCAATGACGTTG 421  
QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
DB 420 TCATTG-----TGGCGGCACCGAGGTCTAGCACAGGGGCTGTGTGCCACAG 370  
QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
DB 369 ACGTGGGGCATCTTGTTCACAGCTCCAGTCTTGGCCAGGGGAGCAGCAGGAGGCC 310  
QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysLeu 94  
DB 309 CTTCCCGCC-----CTGGCTTCACAGATC 286





QY 8 ThrAsnPhaValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg--- 26  
DB 1085 ACACGCTTGGTGGCCCTCCAGGTGACAGCGGCTCCCGCAGGGGAGCATCAGGCC 1026  
QY 27 -----GluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
DB 1025 CCATCCGGGATCTCCACACCATGACTGTGCAGTCTCCGAGCCCACTGCAATGACGTTG 966  
QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
DB 965 TCATTG-----TCCGGGACACAGCGGATGTCTAGCACAGGCGCTGTGGCCACAG 915  
QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
DB 914 ACCGTGGGCGCATCTGTCTCCACACGCTCCAGTCTTCCCGCAGGGCAGCACAGGAAGGCC 855  
QY 75 LysProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
DB 854 CTCTCCCGC-----CTGCGCTCACAGATC 831  
QY 95 ArgGluMetAsn-----LysThrIleSerGlnGluSerAlaArgVa 108  
DB 830 AGGGCCACAACTTAGGTTGACAGCACAGAACGCACTGTCCAGGTGTCTGTGAGACG 771  
QY 108 lAsnHisArgLeuProGluGlyHisPro 117  
DB 770 CGCACATCTTCATAGCACTGTGGCGCT 743  
RESULT 16  
ABQ46392/c  
ID ABQ46392 standard; DNA; 1007 BP.  
XX AC ABQ46392;  
XX DT 12-JUL-2002 (first entry)  
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 32983.  
XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
XX KW drug; side effect; cancer; central nervous system; cardiovascular;  
XX KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
XX KW SNP; cell differentiation; ds.  
XX OS Homo sapiens.  
XX PN WO200218632-A2.  
XX PD 07-MAR-2002.  
XX PF 01-SEP-2001; 2001WO-EPI0074.  
XX PR 01-SEP-2000; 2000DE-1043826.  
XX PR 05-SEP-2000; 2000DE-1044543.  
XX PA (EPIG-) EPIGENOMICS AG.  
XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX WPI; 2002-371829/40.  
XX DR Determining the degree of cytosine methylation in genomic DNA, useful  
XX PT for diagnosis and prognosis, comprises selective hybridization of  
XX PT amplicons from chemically treated DNA  
XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX CC This invention describes a novel method for determining the degree of  
XX CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
XX CC genomic sample of DNA. The sample is treated chemically to convert  
XX CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
XX CC DNA that contains the target C is amplified to form a labeled amplicon.  
XX CC The amplicon is hybridised to two classes, each with at least one  
XX CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers

CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ3410-ABQ411 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.

XX SQ Sequence 1007 BP; 117 A; 140 C; 381 G; 369 T; 0 other;

Alignment Scores:  
Pred. No.: 8.23 Length: 1007  
Score: 87.00 Matches: 29  
Percent Similarity: 47.06% Conservative: 11  
Best Local Similarity: 34.12% Mismatches: 29  
Query Match: 12.29% Indels: 16  
DB: 24 Gaps: 5

US-10-087-573-2 (1-141) x ABQ46392 (1-1007)

QY 54 ProGlyThrGlyAlaSerAlaIleAlaAlaThrValThrProLysGlyAlaSerMetLys 73  
DB 392 CCGTCAACCCCAACACACACCATACGCGCAACCCCTACCCCGCGCTCCCGCAACCCC 333  
QY 74 Leu---LysProProArgProGln-----SerThrLysSerProGluLeuArg--- 88  
DB 332 CTTCGGAACCGCCCGCGCTCCCTAAACCCCGAATTAAATATCCGAAATTCGACGC 273  
QY 89 ---GluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArg 107  
DB 272 TACGAACTCGATACGAAACGCGACGAGAAACATTTTACA-----AAACCGCGAAC 222  
QY 108 ValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGluTyrPheArg 127  
DB 221 GAAATATCACCCTCGACCG-----CGCCAAAAACAAATCTATCAC 183  
QY 128 HisLeuArgSerLeu 132  
DB 182 CGTCTCCGAAACCTTA 168

RESULT 17  
ABQ46393

ID ABQ46393 standard; DNA; 1007 BP.

XX AC ABQ46393;

XX DT 12-JUL-2002 (first entry)

XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 32984.

XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
XX KW drug; side effect; cancer; central nervous system; cardiovascular;  
XX KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
XX KW SNP; cell differentiation; ds.

XX OS Homo sapiens.

XX PN WO200218632-A2.

XX PD 07-MAR-2002.

XX PF 01-SEP-2001; 2001WO-EPI0074.

XX PR 01-SEP-2000; 2000DE-1043826.

XX PR 05-SEP-2000; 2000DE-1044543.

XX PA (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX WPI; 2002-371829/40.  
XX Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA -  
XX  
XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX  
XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ3410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.  
XX  
XX Sequence 1007 BP; 369 A; 381 C; 140 G; 117 T; 0 other;  
SQ  
Alignment Scores:  
Pred. No.: 8.23 Length: 1007  
Score: 87.00 Matches: 29  
Percent Similarity: 47.06% Conservative: 11  
Best Local Similarity: 34.12% Mismatches: 29  
Query Match: 12.29% Indels: 16  
DB: 24 Gaps: 5  
US-10-087-573-2 (1-141) x ABQ46393 (1-1007)  
QY 54 ProGlyThrGlyAlaSerAlaLeAlaAlaThrValThrProLysGlyAlaSerMetLys 73  
Db 616 CGGTCAACCCCAACCAACCAACCACTACGCGCAACCCCTACCCCGCGCTCCCGGAACCCC 675  
QY 74 Leu--LysProArgProGln-----SerThrLysSerProGluLeuArg--- 88  
Db 676 CTTCGCAACCCGCGCGCGCTTAAACCCCGCAATTAATACCCGAATTCGACGC 735  
QY 89 ---GluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArg 107  
Db 736 TACGAACCTCGATACGAAACCGCGGCAAAACCTTTACA-----AAAACCGCGAAC 786  
QY 108 ValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGluTyrPheArg 127  
Db 787 GAAATACCCGTCGACCG-----CGCAAAACAAATCTATCATC 825  
QY 128 HisLeuArgSerLeu 132  
Db 826 CGTCTCCGAACCCCTA 840  
RESULT 18  
ABQ81842/c  
ID ABQ81842 standard; DNA; 349980 BP.  
XX  
XX ABQ81842;  
XX  
XX 19-NOV-2002 (first entry)  
XX  
XX Bifidobacterium longum NCC2705 genomic sequence SEQ ID NO:1.

XX Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;  
KW anti-diarrheic; antibacterial; inhibitor of Salmonella; detection;  
KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;  
XX rotavirus; food composition; pharmaceutical composition; gene; ds.  
XX Bifidobacterium longum.  
XX EP1227152-A1.  
XX 31-JUL-2002.  
XX 30-JAN-2001; 2001EP-0102050.  
XX 30-JAN-2001; 2001EP-0102050.  
XX (NEST) SOC PROD NESTLE SA.  
XX WPI; 2002-668397/72.  
XX Novel polynucleotide comprising Bifidobacterium genome sequence useful  
PT as a probe or primer for detecting and/or identifying Bifidobacterium  
PT longum in a biological sample -  
XX  
XX Claim 1; SEQ ID 1; 80pp; English.  
XX  
XX The present invention describes a polynucleotide (I) comprising a  
CC sequence of a Bifidobacterium genome selected from the nucleotide  
CC sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at  
CC least 90% identity or which hybridises with the sequences given in  
CC ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding  
CC a fusion protein, comprising a sequence selected from 1097 sequences  
CC given in ABP65259 to ABP6354 ligated in frame to a polynucleotide  
CC encoding a heterologous polypeptide. (I) has anti-diarrheic and  
CC antibacterial activities, and can be used as an inhibitor of Salmonella.  
CC (II) (which is a probe) is useful for the detection and/or identification  
CC of Bifidobacterium longum in a biological sample. A carrier containing  
CC the lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM 1-2618)  
CC can be used for preventing and/or treating diarrhoea brought about by  
CC pathogenic bacteria and/or rotavirus. The carrier is a food composition  
CC selected from milk, yogurt, curd, cheese, fermented milks, milk based  
CC fermented products, ice-creams, fermented cereal based products, milk  
CC based powders, infant formula, pet food or a pharmaceutical composition  
CC selected from tablets, liquid bacterial suspensions, dried oral  
CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.  
CC (I) is useful in DNA arrays or chips to carry out analysis of the  
CC expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent  
CC Bifidobacterium related nucleotide sequences given in the Sequence  
CC Listing from the present invention but not mentioned further within the  
CC specification.  
CC N.B. The sequence data for this patent is not represented in the printed  
CC specification but is based on sequence information supplied by the  
CC European Patent Office.  
XX  
XX Sequence 349980 BP; 72540 A; 102738 C; 103221 G; 71481 T; 0 other;  
SQ  
Alignment Scores:  
Pred. No.: 2.63e+04 Length: 349980  
Score: 85.50 Matches: 41  
Percent Similarity: 44.78% Conservative: 19  
Best Local Similarity: 30.60% Mismatches: 50  
Query Match: 12.08% Indels: 24  
DB: 24 Gaps: 6  
US-10-087-573-2 (1-141) x ABQ81842 (1-349980)  
QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
Db 214556 AGCACCAACACCAACGACCGCAATGGCGACCGACCCCGTATTTC---GAAAGGGGC 214500  
QY 23 AspValMetArgGluAlaLeuLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
Db 214499 GCAGCGCCATTAGCAGCGCGCTTATTCTGTCGAGTGTACCGGGCTTAGCAGGTGTCGA 214440

QY 43 AlaLeu-----AlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSer 59  
 Db 214439 GGATTGTGTGATCGCGCGCTTCTGCTCT-----CCAGGGCGGGAGTGACC 214392  
 QY 60 AlaIleAlaIaThrValThrProLysGlyAla-----SerMet 72  
 Db 214391 TCGTGCGCTGACCGTGACTTCGAACGTCGCTTCAGCCAGCAGATAAGAGAGCGGTGATG 214332  
 QY 73 LysLeuLysPro-----ProArgProGlnSerThrLysSerProGluLeuArg 88  
 Db 214331 GTCTCTGCTGCTGCGGTGGCAGAAATCAAGCCAGTCAGATCAGATCAGTTGGA 214272  
 QY 89 GluLeuSerArgLysIleArgGluMetAsnLysThrIle-----SerGlnGlu 104  
 Db 214271 TCAATCAGCTTGCAGCGGTTTCAGTATGCCGACITTTGAATAGCGGATACCTTCACA 214212  
 QY 105 SerAlaArgValAsn-HisArgLeuProGluGlyHisPro 117  
 Db 214211 TCGGCAGCGTGAACCTATCGCTAGCTGATAGGAATCCT 214172  
 RESULT 19  
 AAF13411  
 ID AAF13411 standard; cDNA; 1337 BP.  
 AC AAF13411;  
 DT 13-MAR-2001 (first entry)  
 DE Aspergillus oryzae EST SEQ ID NO:5934.  
 KW Multiple gene expression; filamentous fungal cell; EST;  
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
 KW culture condition; environmental stress; spore morphogenesis;  
 KW metabolic pathway engineering; catabolic pathway engineering; ss.  
 OS Aspergillus oryzae.  
 XX WO200056762-A2.  
 XX 28-SEP-2000.  
 XX 22-MAR-2000; 2000WO-US07781.  
 XX 22-MAR-1999; 99US-0273623.  
 XX (NOVO ) NOVO NORDISK BIOTECH INC.  
 XX (NOVO ) NOVO NORDISK AS.  
 XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
 WPI; 2000-594572/56.  
 PT Monitoring differential expression of genes in filamentous fungal cells  
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
 PT substrate of expressed sequence tags -  
 XX Claim 88; Page 2451; 3161pp; English.  
 XX The present invention describes a method for monitoring differential  
 XX expression of genes in a first filamentous fungal (FF) cell relative to  
 XX expression of the same genes in one or more second filamentous fungal  
 XX cells. The method uses fluorescence-labeled nucleic acids isolated from  
 XX the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 XX are used in the methods for monitoring differential expression of genes  
 XX in a first filamentous fungal (FF) cell relative to expression of the  
 XX same genes in one or more second filamentous fungal cells. Monitoring  
 XX the global expression of genes from FF cells allows the production  
 XX potential of the microorganisms to be improved. New genes may be  
 XX discovered, possible functions of unknown open reading frames can be  
 XX identified and gene copy number variation and stability can be  
 XX monitored. The expression of genes can be used to study how FF cells

CC adapt to changes in culture conditions, environmental stress, spore  
 CC morphogenesis, recombination, metabolic or catabolic pathway  
 CC engineering. Using ESTs provides several advantages over genomic or  
 CC random cDNA clones including elimination of redundancy as one spot on an  
 CC array equals one gene or open reading frame, and organisation of the  
 CC microarrays based on function of the gene products to facilitate  
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from  
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus  
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and  
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are  
 CC all specifically claimed in the present invention.  
 XX  
 SQ Sequence 1337 BP; 368 A; 318 C; 310 G; 341 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 19.5 Length: 1337  
 Score: 85.00 Matches: 35  
 Percent Similarity: 42.25% Conservative: 25  
 Best Local Similarity: 24.65% Mismatches: 42  
 Query Match: 12.01% Indels: 40  
 DB: 21 Gaps: 7  
 US-10-087-573-2 (1-141) x AAF13411 (1-1337)  
 QY 2 GluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThr 21  
 Db 273 GAGCAACCGTGATCAAAAGCTACTCAACAGCAGAGCGCCAGCCA----- 317  
 QY 22 PheAspValMetArgGluAlaLeuLeuArgVallys-----SerSerGluArg----- 37  
 Db 318 TATGACATCGTACGCGCAGCGCGCTCAAGAGGTTGAACCTGGATAGCTCTCGCCGGTGAGAA 377  
 QY 38 -----LeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHis 50  
 Db 378 CTTGAAGACTACGTGGATCTTATCTCCAAGCTGCGCTACTTCCGGGCAAGCGGG--- 434  
 QY 51 ArgValLeuProGlyThrGlyAlaSerAlaIleAlaIaThrValThrProLysGlyAla 70  
 Db 435 -----CCTACCGGCACCTGCCAGTCCCGCGGTAGTACCGGTAATGCGCGCT 482  
 QY 71 SerMetLysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeu 90  
 Db 483 GGA-----TCATCTGCCCATGTCTCAAAGGCGAGCATA 515  
 QY 91 SerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHis 110  
 Db 516 GCGGAAGAAATTCAC-----GTCCAGGGTTTCATCGGCAAAATGTAATCAT 560  
 QY 111 ArgLeuProGlyGlyHisProLeuLeuGluLysArgAlaGluTyrPheArgHisLeuArg 130  
 Db 561 ACTATCAATGAG-----GATGAAGGACCGAGTTCACGAGACACATCAAT 605  
 QY 131 SerLeu 132  
 Db 606 GCCGTT 611  
 RESULT 20  
 AAN60288  
 ID AAN60288 standard; DNA; 9213 BP.  
 XX  
 AC AAN60288;  
 XX  
 DT 08-JUN-1991 (first entry)  
 XX  
 DE Sequence of the HTLV-III genome.  
 XX  
 KW HIV; LAV; AIDS; diagnosis; vaccine; ss.  
 XX  
 OS HTLV-III/H9 cells (ATCC CRL 8543).  
 XX  
 FH Key Location/Qualifiers  
 FT repeat\_region 1..96  
 FT /\*tag= a

US-10-087-573-2 (1-141) x AAN60288 (1-9213)

QY	22	PheAspValMetArgGluAlaLeuLeuArgVallySerSerGluArgLeuAlaMetLeu	41
Db	4955	TTTGACTGTTTTTCAGACTCTGCTATAAGA	4984
QY	42	ArgAlaLeuAlaGly	52
Db	4985	AGGGCCTTATTAGGACACATAGTTAGCCCTAGCTGTGAATATCAAGCAGGACATACAAAG	5044
QY	53	LeuProGlyThrGlyAlaSerAlaIleAlaIleAlaThrValThrProLysGlyAlaSerMet	72
Db	5045	GTAGGATCTCTACAATACATTGGCACTAGGCAGCATTAATAACACCAAAA	5092
QY	73	LysLeuLysProProArgProGlnSerThrIysSerProGluLeuArgGluLeuSerArg	92
Db	5093	AAAGATAAGCCACCTTTGGCTAGTTAGCGAACTGACAGAGGATAGA	5140
QY	93	LysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeu	112
Db	5141	-----TGGACAGCCCCAGAGACCAGAGGCCACAGAGGAGCCACAGAAATC	5188
QY	113	ProGluGlyHis--ProLeuLeuGluLys--ArgAlaGluTyrPheArgHisLeuArgS	131
Db	5189	---AATGGACACTAGAGCTTTTAGAGGAGCTTAAGAATGAAGCTGTAGACATTTTCCTA	5245
QY	131	erLeuLysSerGlnGlyValAsnArgLeuIle	141
Db	5246	GGATTGGCTCCATGCTTAGGGCAACATATC	5277

RESULT 21  
ABV21005/c  
ID ABV21005 standard; cDNA; 1972 BP.  
AC ABV21005;  
XX  
XX 13-SEP-2002 (first entry)  
XX Human prostate expression marker cDNA 20996.  
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
XX pharmacogenomic marker; gene; ss.  
XX Homo sapiens.  
XX WO200160860-A2.  
XX  
XX 23-AUG-2001.  
XX  
XX 20-FEB-2001; 2001WO-US05171.  
XX  
XX 17-FEB-2000; 2000US-183319P.  
XX PR 16-MAR-2000; 2000US-189862P.  
XX PR 25-MAY-2000; 2000US-207454P.  
XX PR 09-JUN-2000; 2000US-211314P.  
XX PR 18-JUL-2000; 2000US-219007P.  
XX PR 13-DEC-2000; 2000US-255281P.  
XX  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX PA  
XX Schlegel R, Endege WO, Monahan JE;  
XX  
XX WPI; 2001-662795/76.  
XX  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
XX prostate cells and correlating with presence of prostate cancer, useful  
XX for detecting presence of prostate cancer, stage of prostate cancer -  
XX  
XX Claim 1; Page 3457; 11750pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
XX a nucleotide sequence given in Tables 1-9 (ABV0010-ABV62213) of the  
XX CC specification or its complement. (I) is useful for:

CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 1972 BP; 490 A; 472 C; 557 G; 452 T; 1 other;  
  
Alignment Scores:  
Pred. No.: 53.2 Length: 1972  
Score: 83.00 Matches: 35  
Percent Similarity: 43.01% Conservative: 5  
Best Local Similarity: 37.63% Mismatches: 27  
Query Match: 11.72% Indels: 26  
DB: 23 Gaps: 4  
  
US-10-087-573-2 (1-141) x ABV21005 (1-1972)  
QY 31 ArgValLysSerSerGluArgLeuAlaMet-LeuArgAlaLeuAlaGlyMetCysGlyHi 50  
Db 362 CGCACGCTCAGCTCTCCCAAGTTGGTTCCCAACCGCGCACTCGGAGTATGCGACATT 303  
QY 50 sArgValLeu-----ProGlyThrGlyAlaSerAlaAlaAlaThrVa 65  
Db 302 CCGAGTTGTCTATCATACAGTGGCGGCGAGCCCGGGCTCTCTCTCGCGGAGCCATGTT 243  
QY 65 lThrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSerThrLysSerPr 85  
Db 242 GCGCGCGCTCGAGGCC-----CGCAGCCCGCGGCC---GGAACCCGAGAGCCC 198  
QY 85 oGlu-LeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGlu 105  
Db 197 AGAGCTTCGACGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 175  
QY 105 erAlaArgValAsnHisArgLeuProGluGlyHis 116  
Db 174 -----CGGGCTGGCGGCCGGAACCCGAGGCCAC 145  
  
RESULT 22  
ABV26849/C  
ID ABV26849 standard; cDNA; 1972 BP.  
XX AC ABV26849;  
XX DT 16-SEP-2002 (first entry)  
XX DE Human prostate expression marker cDNA 26840.  
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX OS Homo sapiens.  
XX PN WO200160860-A2.  
XX DT 23-AUG-2001.  
XX PF 20-FEB-2001; 2001WO-US05171.  
XX PR 17-FEB-2000; 2000US-183319P.  
XX PR 16-MAR-2000; 2000US-189862P.  
XX PR 25-MAY-2000; 2000US-207454P.  
XX PR 09-JUN-2000; 2000US-211314P.  
XX PR 18-JUL-2000; 2000US-219007P.  
XX PR 13-DEC-2000; 2000US-255281P.  
XX

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX Schlegel R, Endege WO, Monahan JE;  
PI WPI; 2001-662795/76.  
DR  
XX  
PT Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PI for detecting presence of prostate cancer, stage of prostate cancer  
XX  
PS Claim 1; Page 5426; 11750pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 1972 BP; 490 A; 472 C; 557 G; 452 T; 1 other;  
  
Alignment Scores:  
Pred. No.: 53.2 Length: 1972  
Score: 83.00 Matches: 35  
Percent Similarity: 43.01% Conservative: 5  
Best Local Similarity: 37.63% Mismatches: 27  
Query Match: 11.72% Indels: 26  
DB: 23 Gaps: 4  
  
US-10-087-573-2 (1-141) x ABV26849 (1-1972)  
QY 31 ArgValLysSerSerGluArgLeuAlaMet-LeuArgAlaLeuAlaGlyMetCysGlyHi 50  
Db 362 CGCACGCTCAGCTCTCCCAAGTTGGTTCCCAACCGCGCACTCGGAGTATGCGACATT 303  
QY 50 sArgValLeu-----ProGlyThrGlyAlaSerAlaAlaAlaThrVa 65  
Db 302 CCGAGTTGTCTATCATACAGTGGCGGCGAGCCCGGGCTCTCTCTCGCGGAGCCATGTT 243  
QY 65 lThrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSerThrLysSerPr 85  
Db 242 GCGCGCGCTCGAGGCC-----CGCAGCCCGCGGCC---GGAACCCGAGAGCCC 198  
QY 85 oGlu-LeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGlu 105  
Db 197 AGAGCTTCGACGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 175  
QY 105 erAlaArgValAsnHisArgLeuProGluGlyHis 116  
Db 174 -----CGGGCTGGCGGCCGGAACCCGAGGCCAC 145  
  
RESULT 23  
AAS00717  
ID AAS00717 standard; DNA; 1812 BP.  
XX AC AAS00717;  
XX DT 12-SEP-2001 (first entry)  
XX DE Chimeric Caufaq DNA polymerase version 1 polynucleotide.  
XX KW Caufaq DNA polymerase; 3'-5' exonuclease; mutant; proof-reading;  
KW second strand DNA synthesis; PCR; ss.  
XX

OS Chimeric - Thermus aquaticus.  
XX Chimeric - Unidentified.  
FH Key Location/Qualifiers  
FT 1..1812  
FT /\*tag= a  
FT /product= "Chimeric Cautaq DNA polymerase"  
FT /partial  
FT /note= "No stop codon present"  
XX  
PN WO200118213-A1.  
PP 15-MAR-2001.  
PP 08-SEP-2000; 2000WO-GB03478.  
PP 09-SEP-1999; 99GB-0021318.  
PP (DZIE/) DZIEGLEWSKA H.  
PP (KRIS/) KRISTENSEN T.  
PI Kristensen T;  
XX WPI; 2001-218561/22.  
DR P-PSDB; AAU00575.  
XX Chimeric DNA polymerase useful in molecular biology and gene technology  
PT techniques, has domain having proof-reading activity from a DNA  
PT polymerase and DNA polymerase domain from heterologous DNA polymerase -  
XX Claim 13; Page 44-48; 66pp; English.  
CC The sequence represents a polynucleotide encoding a chimeric Cautaq DNA  
CC polymerase polypeptide. The chimeric polypeptide comprises a 3'-5'  
CC exonuclease domain from one DNA polymerase and a DNA polymerase domain  
CC from a heterologous DNA polymerase. This molecule has a domain exhibiting  
CC 3'-5' exonuclease activity derived from Cau DNA polymerase and another  
CC exhibiting DNA polymerase activity from Taq DNA polymerase. These  
CC chimeric sequences are useful in molecular biology and gene technology  
CC techniques, especially second strand DNA synthesis, PCR amplification and  
CC DNA sequencing. The proof-reading (3'-5'exonuclease) activity makes the  
CC polymerase particularly useful in techniques where the fit between the  
CC primer and the template DNA is not exact.  
XX  
SQ Sequence 1812 BP; 334 A; 574 C; 595 G; 319 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 53.8 Length: 1812  
Score: 82.50 Matches: 47  
Percent Similarity: 38.01% Conservative: 18  
Best Local Similarity: 27.49% Mismatches: 47  
Query Match: 11.65% Indels: 59  
DB: 22 Gaps: 8  
  
US-10-087-573-2 (1-141) x AAS00717 (1-1812)  
QY 24 ValMetArgGluAlaLeuLeuArgVallySerSerGluArgLeuAlaMetLeuArgAla 43  
DB 470 ATAAGCAGATTAGCTTGTCTGAGGTGCGGTGAGCAGCGGCTCCGTATGCGGCGCG 529  
QY 44 Leu---AlaGlyMetCysGly-HisArgValLeu----- 53  
DB 530 ATGCCCTCTACACGTTCGGGCTCACCAGCGCTTCAGCGCTGAGGCTGAACCTG 589  
QY 53 ----- 53  
DB 590 CGCTGCAGATCTGTATTACAGAGTGAGAGGCCCTTTTCGGTGTCTGCGCCACATGG 649  
QY 54 -----ProGlyThrGlyAlaSerAlaIleAlaIleValThrProLy 68  
DB 650 AGGCCACGGGGGTGGCGCTGACGTGGC-----CTATCTCAGGCGCTTGTCTCT-- 698  
QY 68 sGlyAlaSerMetLysLeuLysProProArg----- 78

Db 699 -GGAGGTGGCGGAGGATCGCCGCGCTCGAGGCGCGAGGTCTTCCGCTGGCGGCCACC 757  
QY 79 -----ProGln-----SerThrLysSerProGluLeuArgGluLeuSerAr 92  
Db 758 CATTCAACCTCAACTCCCGGACAGCTGGAAGGGTCTCTTTGACGAGCTAGGGCTTC 817  
QY 92 gLysIleArgGlu-MetAsnLysThrIleSerGlnGlu---SerAlaArgValAsnHisA 111  
Db 818 CCGCCATCGGCAAGACGGAGAGACCGGCAAGCGCTCCACACGCGCGCTCTCTGGAGG 877  
QY 111 rgLeuProGluGlyHisProLeuLeuGluLysArgAlaGluTyrPheArgHisLeuArgS 131  
Db 878 CCTCCGCGAGGCCACCCCATCTGTGAGAAGATCTCTGCAGTAC---CGGAGGCTCACCA 934  
QY 131 erLeuLysSerGlnGlyValAsnArgLeu 140  
Db 935 AGCTGAAGAGCACCTACATTGACCCCTTG 963  
  
RESULT 24  
AAS00718  
ID AAS00718 standard; DNA; 1812 BP.  
XX AC  
XX AAS00718;  
XX  
DT 12-SEP-2001 (first entry)  
DE Chimeric Cautaq DNA polymerase version 2 polynucleotide.  
XX  
XX Cautaq DNA polymerase; 3'-5' exonuclease; mutant; proof-reading;  
XX second strand DNA synthesis; PCR; ss.  
OS Chimeric - Thermus aquaticus.  
OS Chimeric - Unidentified.  
FH Key Location/Qualifiers  
FT CDS 1..1812  
FT /\*tag= a  
FT /product= "Chimeric Cautaq DNA polymerase"  
FT /partial  
FT /note= "No stop codon present"  
PN WO200118213-A1.  
XX  
XX 15-MAR-2001.  
XX 08-SEP-2000; 2000WO-GB03478.  
XX 09-SEP-1999; 99GB-0021318.  
XX (DZIE/) DZIEGLEWSKA H.  
XX (KRIS/) KRISTENSEN T.  
PI Kristensen T;  
XX WPI; 2001-218561/22.  
DR P-PSDB; AAU00574.  
XX Chimeric DNA polymerase useful in molecular biology and gene technology  
PT techniques, has domain having proof-reading activity from a DNA  
PT polymerase and DNA polymerase domain from heterologous DNA polymerase -  
XX Claim 13; Page 52-56; 66pp; English.  
CC The sequence represents a polynucleotide encoding a chimeric Cautaq DNA  
CC polymerase polypeptide. The chimeric polypeptide comprises a 3'-5'  
CC exonuclease domain from one DNA polymerase and a DNA polymerase domain  
CC from a heterologous DNA polymerase. This molecule has a domain exhibiting  
CC 3'-5' exonuclease activity derived from Cau DNA polymerase and another  
CC exhibiting DNA polymerase activity from Taq DNA polymerase. These  
CC chimeric sequences are useful in molecular biology and gene technology  
CC techniques, especially second strand DNA synthesis, PCR amplification and  
CC DNA sequencing. The proof-reading (3'-5'exonuclease) activity makes the  
CC polymerase particularly useful in techniques where the fit between the  
CC primer and the template DNA is not exact.  
XX  
SQ Sequence 1812 BP; 334 A; 574 C; 595 G; 319 T; 0 other;



239 GAGCGAGGCCCTTGAAGCGGAGATACCTCTGACTAGGAGCACTCCCAAGACCTGGAGG 180  
103 GlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArg 122  
179 -----GAGCTACGGCTGGCTTTGAAGCCCGCAAGATACCCACATTA 138  
123 AlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeuLeu 141  
137 ACGCGGCATTAGACCCCTTAACACACTATCCCTTTACCCCTATCAATGAACCTCTTG 81

RESULT 26  
AAS69040  
ID AAS69040 standard; cDNA; 597 BP.  
XX  
AC AAS69040;  
XX

13-FEB-2002 (first entry)  
XX  
XX DNA encoding novel human diagnostic protein #4844.  
XX  
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX  
KW human supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX

WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
30-MAR-2001; 2001WO-US08631.  
XX  
31-MAR-2000; 2000US-0540217.  
PR  
23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSEQ-) HYSEQ INC.  
PA  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
XX  
XX P-PSDB; ABG04853.  
XX

New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 1; SEQ ID No 4844; 103pp; English.  
XX

The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (II) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 597 BP; 97 A; 217 C; 196 G; 87 T; 0 other;  
XX  
Alignment Scores:

Pred. No.: 14.1 Length: 597  
Score: 82.00 Matches: 32  
Percent Similarity: 40.95% Conservative: 11  
Best Local Similarity: 30.48% Mismatches: 38  
Query Match: 11.58% Indels: 24  
DB: Gaps: 3  
US-10-087-573-2 (1-141) x AAS69040 (1-597)  
QY 37 ArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgVal-----LeuPro 54  
Db 213 CGCCTCACACCGCTTCGTGCAGCGCACCGCCTGCTGGCCACCGCTGGGAGACGTCCA 272  
QY 55 GlyThrGlyAlaSerAlaIleAlaAlaThrValThrProLysGlyVala----- 70  
Db 273 GGAGGCAGATCCTCTGCTCTCCCTGCGGTGCTACAGGAAGTGGAGCTAAGGCTGCTGG 332  
QY 71 -----SerMetLysLeuLysProProArgProGlnSerThrLysSerProGlu 86  
Db 333 CGAGGCCACCTGTCAATGCTCTACAGCAGCCCGCTCCCTCAACCTCCTCCAGAT 392  
QY 87 LeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAla 106  
Db 393 -----ATTGCCAGGATGCTGTG 410  
QY 107 ArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGluTyrPhe 126  
Db 411 TGCTGGCTACCCAGAGCGCGCAGGAGACACATGCCAGCCAGTCCAGAGCCCGGTCCAAT 470  
QY 127 ArgHisLeuArgSer 131  
Db 471 GTGCATCAGCGAAGC 485

RESULT 27  
AAS32623  
ID AAS32623 standard; DNA; 10646 BP.  
XX  
AC AAS32623;  
XX  
DT 17-DEC-2001 (first entry)  
XX  
XX Human genomic DNA for novel endocrine antigen, SEQ ID No 577.  
DE Human; endocrine antigen; ds; cytostatic; antiinfertility; antidiabetic;  
XX thyroid-active; adrenal-active; androgenic; gastric; gene therapy;  
KW antisense-therapy; antibody; endocrine disorder; hormone imbalance;  
KW reproductive disorder; endocrine cancer; pancreatic disorder;  
KW diabetes mellitus; adrenal gland disorder; hirsutism; thyroid disorder;  
KW hyperthyroidism; hypothalamic disorder; vanishing testes syndrome.  
XX  
OS Homo sapiens.  
XX  
PN WO200155319-A2.  
XX  
PD 02-AUG-2001.  
XX

17-JAN-2001; 2001WO-US01335.  
XX  
31-JAN-2000; 2000US-0179065.  
PR  
04-FEB-2000; 2000US-0180628.  
PR  
24-FEB-2000; 2000US-0184664.  
PR  
02-MAR-2000; 2000US-0186350.  
PR  
16-MAR-2000; 2000US-0189874.  
PR  
17-MAR-2000; 2000US-0190076.  
PR  
18-APR-2000; 2000US-0198123.  
PR  
19-MAY-2000; 2000US-0205515.  
PR  
07-JUN-2000; 2000US-0209467.  
PR  
28-JUN-2000; 2000US-0214886.  
PR  
30-JUN-2000; 2000US-0215135.  
PR  
07-JUL-2000; 2000US-0216647.  
PR  
11-JUL-2000; 2000US-0216880.  
PR  
11-JUL-2000; 2000US-0217487.  
PR  
11-JUL-2000; 2000US-0217496.  
PR



PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226868.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236902.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249267.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-02559678.  
 XX

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-457726/49.

Isolated polypeptide for treating, preventing and prognosing disorders related to the endocrine system including endocrine disorders, reproductive disorders, and gastrointestinal disorders and also for testing and detection e.g. diagnosis -

Disclosure; SEQ ID No 577; 558pp; English.

The invention relates to cDNAs encoding novel human endocrine antigens or a fragment having biological activity, a domain, an epitope, full length protein, variant, allelic variant or a species homologue of the cDNA/antigen. The DNAs and polypeptides are useful for preventing, treating or ameliorating a medical condition when administered (e.g. by gene therapy or antisense-therapy). Identifying mutations in the genes coding for the antigens is useful for diagnosing a pathological condition or a susceptibility to a pathological condition. The DNAs, antigens and antibodies raised against the antigens useful for treating, preventing and/or prognosing disorders related to the endocrine system or hormone imbalance or reproductive disorders, cancers of endocrine

CC tissues, disorders of the pancreas (e.g. diabetes mellitus), the adrenal  
CC glands (e.g. hirsutism), ovaries, the thyroid (e.g. hyperthyroidism), the  
CC hypothalamus and testes (e.g. vanishing testes syndrome), many examples  
CC of diseases and disorders are given in the specification. The present  
CC sequence is genomic DNA fragment form a gene encoding an endocrine  
CC antigen of the invention  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 10646 BP; 2501 A; 3544 C; 2986 G; 1615 T; 0 other;

Alignment Scores:  
Pred. No.: 626 Length: 10646  
Score: 82.00 Matches: 23  
Percent Similarity: 45.95% Conservative: 11  
Best Local Similarity: 31.08% Mismatches: 32  
Query Match: 11.58% Indels: 8  
DB: 22 Gaps: 2

US-10-087-573-2 (1-141) x AAS32623 (1-10646)

Qy 46 GlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAlaIleAlaAlaThrVal 65  
Db 5337 GGGGCTACGGGACCGGGTCTGAGTGGGACAGGCTGAGGTTTGCAACAGGTCACGCTG 5396  
Qy 66 ThrProLysGlyAlaSerMetLysLeuLysProProArgProGln-----Ser 81  
Db 5397 CCACGAGCAGACAGACAGGCTCATGTTCCCATGAAGCCCCAAAGCAGAGAGCTCCAGC 5456  
Qy 82 ThrLysSerProGluLeuArgGlu-----LeuSerArgLysIleArgGluMet 97  
Db 5457 ACAGCAGCTCCACTTCGACGAGGAGGAGAGGACAGCCAGGCGGCTTCAAGTTCTCTG 5516  
Qy 98 AsnLysThrIleSerGlnGluSerAlaArgValAsnHisArg 111  
Db 5517 CTCTCCAGCTGCAGCAGCAGCAGCTCAGAGAGGCGCCACAGG 5558

RESULT 28  
AAS06738  
ID AAS06738 standard; cDNA; 2838 BP.  
XX AAS06738;  
AC AAS06738;  
DT 12-SEP-2001 (first entry)  
XX Polynucleotide sequence encoding human protein kinase #38.  
DE Human; protein kinase; PTK; STK; cancer; cardiovascular disease;  
KW metabolic disorder; immune related disease; neurological disorder;  
KW neurodegenerative disorder; inflammatory disorder; infectious disease;  
KW reproductive disorder; gene therapy; ss.  
XX Homo sapiens.  
OS

XX WO200138503-A2.  
PN 31-MAY-2001.  
PD 22-NOV-2000; 2000WO-US32085.  
PF 24-NOV-1999; 99US-0167482.  
PR (SUGEN-) SUGEN INC.  
PA Plowman GD, Whyte D, Manning G, Sudarsanam S, Martinez R,  
PI Flanagan P, Clary D;  
FI WPI; 2001-343950/36.  
DR P-PSDB; AAU03538.  
DR Nucleic acids encoding human kinase polypeptides, useful for preventing

PT diagnosing and/or treating e.g. cancer, immune, cardiovascular and  
PT neuronal-associated diseases, and microbial infections -  
XX Example 1; Figure 1; 433pp; English.

XX AAS06701-AAS06757 encode for novel human protein kinases #1-57. The  
CC novel protein kinases have been identified as members of the tyrosine  
CC or serine/threonine kinase (PTK and STK) families. The polynucleotides  
CC encoding protein kinases and the polypeptides may be used in the  
CC prevention, diagnosis and treatment of diseases associated with  
CC cancers (especially cancers of haematopoietic origin), cardiovascular  
CC disease (e.g. atherosclerosis), metabolic disorders (e.g. diabetes),  
CC immune related diseases (e.g. rheumatoid arthritis), neurological  
CC disorders (e.g. schizophrenia), neurodegenerative disorders (e.g.  
CC Parkinson's disease), inflammatory disorders (e.g. asthma), infectious  
CC disease (e.g. HIV) and reproductive disorders (e.g. infertility).  
CC Additionally, polynucleotides encoding protein kinases may be  
CC used for gene therapy and as DNA probes in diagnostic assays.  
CC The protein kinase polypeptides may be used as antigens in the production  
CC of antibodies against the protein kinases and in assays to identify  
CC modulators of protein kinase expression and activity.

XX Sequence 2838 BP; 694 A; 934 C; 713 G; 497 T; 0 other;

Alignment Scores:  
Pred. No.: 124 Length: 2838  
Score: 81.50 Matches: 30  
Percent Similarity: 41.51% Conservative: 14  
Best Local Similarity: 28.30% Mismatches: 37  
Query Match: 11.51% Indels: 25  
DB: 22 Gaps: 5

US-10-087-573-2 (1-141) x AAS06738 (1-2838)

Qy 43 AlaLeuAlaGlyMetCys-----GlyHisArgValLeuProGly 55  
Db 676 GCCTACCCGGGCTGTGTCGCCGCCGCCGACCTGAGTCCGGCCACCGTTCCCTGCCCCCA 735  
Qy 56 ThrGlyAlaSerAla-----IleAlaAlaThr 64  
Db 736 TCGCCCCGGCAGCGGACGCGGTCCGCACCCCGCGGACCCCAACATGTCACACC 795  
Qy 65 ValThrProLysGlyAla-----SerMetLysLeuLysProProArgPro 79  
Db 796 GTGACCCCGCGGCGACGCGGCCCATGAGGAAGAAGAACAGCTGAAGCCCGCGGACC 855  
Qy 80 GlnSerThrLysSerProGluLeuArgGluSerArgLysIleArgGluMetAsnLys 99  
Db 856 CCACCGCCCTCTCCGAAAACATGATACACTTGCATCCCGGATTCACCGGCTGCATCGG 915  
Qy 100 ThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHis---ProLeu 118  
Db 916 AGCAAAATCCACGAG---TTCCAGCTGGGGCACCGGCTGGAGCGAGGCCACACGCCCAAA 972  
Qy 119 LeuGluLysArgAlaGlu 124  
Db 973 GCCAAGAGAAGACAAA 990

RESULT 29  
AAD30562  
ID AAD30562 standard; cDNA; 2962 BP.  
XX AAD30562;  
AC AAD30562;  
DT 21-MAY-2002 (first entry)  
XX Human kinase polypeptide (PKIN-15) cDNA.  
DE Human; kinase polypeptide; PKIN-15; gene therapy; Addison's disease;  
KW leukaemia; immune disorder; lymphoma; melanoma; developmental disorder;  
KW acquired immunodeficiency syndrome; AIDS; allergy; anaemia; hypertension;  
KW asthma; Crohn's disease; rheumatoid arthritis; bursitis; atherosclerosis;



XX The present invention relates to methods of detecting a prostate  
CC cancer-associated transcript in a cell from a patient. The method  
CC comprises contacting a biological sample from the patient with  
CC prostate cancer-associated polynucleotides (designated PC genes) that  
CC selectively hybridize to a sequence that is at least 80% identical  
CC to them. The prostate cancer-associated polynucleotide sequences  
CC are differentially expressed in prostate tumour tissue or in  
CC prostate cancer and are derived from the tissues of various  
CC organisms such as humans or other mammals (e.g. mice, sheep and dogs).  
CC The methods of the invention are useful for diagnosing and treating  
CC prostate cancer in mammals. The prostate cancer-associated genes are  
CC useful for diagnosing or treating prostate cancer, as well as for  
CC identifying modulators of prostate cancer or agents that inhibit  
CC prostate cancer. The nucleic acid sequences are particularly useful  
CC in gene therapy, as a vaccine or in antisense applications.  
CC ABK92115-ABK92263 represent prostate cancer-associated polynucleotide  
CC sequences.  
XX  
SQ Sequence 3793 BP; 1269 A; 776 C; 917 G; 831 T; 0 other;

Alignment Scores:  
Pred. No.: 182 Length: 3793  
Score: 81.50 Matches: 37  
Percent Similarity: 39.13% Conservative: 17  
Best Local Similarity: 26.81% Mismatches: 43  
Query Match: 11.51% Indels: 41  
DB: 24 Gaps: 6

US-10-087-573-2 (1-141) x ABK92234 (1-3793)

QY 20 GluThrPheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAla 39  
DB 1963 GAACAATTGGATGCCATCAACAAGAAATCAGGCTAATTCAGGAAGAAAAATCTACA 2022  
QY 40 MetLeuAlaGala-----LeuAlaGlyMetCys 48  
DB 2023 GAGTTGCGTGCTGAGAAATTTGAAATAGAGTGGCTAGTGAGCTCGAAGGCTGAAAT 2082  
QY 49 GlyHisArgValLeuProGlyThrGly-----AlaSerAlaIleAla 62  
DB 2083 TTGGCAAGGGTCCACCAGGTACCTCCATTACTGCTCTGTTACAGCTTCATCGTGCC 2142  
QY 63 AlaThrValThrProLysGlyAlaSerMet---LysLeuLysProProArgProGln--- 80  
DB 2143 AGTTTCATCTCCCTCCAGTGGACACTCAACTCCAAAGCTCACCCCTCGAAGCCCTGCCAGG 2202  
QY 81 -----SerThrLysSerProGluLeuArgGluLeuSerArgLys 93  
DB 2203 GAAATGGATCGGATCGGAGTCATGACACTGCCAAGTGTATCTGAGGAACATCGGAGAAAG 2262  
QY 94 Ile-----ArgGluMetAsnLysThrIleSerGlnGlnSerAla 106  
DB 2263 ATTGCAGTTGTGGAAGAAGATGGTCGAGAGGACAAACAAATTAATGTCAAACTTCT 2322  
QY 107 -----ArgValAsnHisArgLeuProGluGlyHis 116  
DB 2323 CCTCTCTCTACCCCTAGAGCCCTCAGATGACTACACTCTCCCTTCTTCCTAC 2376

RESULT 31  
ID AAT14180  
XX AAT14180 standard; cDNA; 9718 BP.  
XX AAT14180;  
XX  
DT 03-JUL-1996 (first entry)  
XX Attenuated HIV-1 strain CpG1 genome.  
XX HIV-1; human immunodeficiency virus type 1; AIDS; attenuation;  
KW vaccine; DNA methylation; ss.  
XX

OS Synthetic.  
XX WO9611280-A1.  
FN 18-APR-1996.  
XX  
PD 05-OCT-1995; 95WO-US13219.  
XX  
PF 07-OCT-1994; 94US-0319974.  
XX  
PR (UYEC-) UNIV EAST CAROLINA.  
XX  
PA Nyce JW;  
XX WPI; 1996-209861/21.  
XX  
DR Attenuated virus contg. additional methylation sites in its genome  
XX - used to induce protective immunity and to raise antibodies for  
PT diagnostic use.  
PT  
XX  
PS Example 1; Page 23-28; 40pp; English.  
XX  
CC A novel, attenuated HIV-1 genome, HIV-1 CpG1 (AAT14180), is  
CC hypersubstituted with noninformational or 'silent' CpG segments.  
CC Addition of the CpGs to the genome does not alter the amino acid  
CC sequence of the encoded proteins but makes the synthetic genome a  
CC target for host cell-methylases. Thus, although the virus for which  
CC this genome codes is capable of infecting a cell, the proviral genome  
CC is easily inactivated by methylation and kept permanently in a dormant  
CC state. The virus can be used for the prepn. of live virus vaccine or  
CC to raise diagnostic antibodies.  
XX  
SQ Sequence 9718 BP; 2797 A; 2198 C; 2771 G; 1952 T; 0 other;

Alignment Scores:  
Pred. No.: 628 Length: 9718  
Score: 81.50 Matches: 34  
Percent Similarity: 37.50% Conservative: 11  
Best Local Similarity: 28.33% Mismatches: 38  
Query Match: 11.51% Indels: 37  
DB: 17 Gaps: 6

US-10-087-573-2 (1-141) x AAT14180 (1-9718)

QY 22 PheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeu 41  
DB 5373 TTCGACTGTTTTTCGGACTCGGCGATACGC-----5402  
QY 42 ArgAlaLeuAlaGly-----MetCysGlyHisArgVal 52  
DB 5403 AAGCGGTACTCGGACACATCGTTTCGCCCGCTGCGAATATCAAGCGGACATAACAAG 5462  
QY 53 LeuProGlyThrGlyAlaSerAlaIleAlaAlaThrValThrProLysGlyAlaSerMet 72  
DB 5463 GTCGGATCGGTACAAATACCTCGCGCTCGCGCGCTTAATACGCGGAAA-----5510  
QY 73 LysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArg 92  
DB 5511 AAGATAAAGCGCGCGTTTCGCGTTCGAAACTGACGAGGATCGA-----5558  
QY 93 LysIleArgGluMetAsnLysThrIleSerGlnGlnSerAlaArgValAsnHisArgLeu 112  
DB 5559 -----TGGAAACAAGCCCCAGACAGACAGGCGCCACAGAGGCGGACCAATG 5606  
QY 113 ProGluGlyHis-----ProLeuLeuGluLysArgAlaGluTyrPheArg 127  
DB 5607 ---AATGCACACTAGAGCTTTTAGAGGAGCTTAAAGAACGAAGCGGTTCGCCATTTTCGCGC 5663  
RESULT 32  
ABL28661/c  
ID ABL28661 standard; DNA; 1126 BP.  
XX  
AC ABL28661;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 37456.

XX KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical; gene; ds.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX PI WPI; 2001-656860/75.

XX DR New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

XX PT Claim 1; SEQ ID NO 37456; 21pp + Sequence Listing; English.

XX PS The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

XX CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1126 BP; 263 A; 320 C; 340 G; 203 T; 0 other;

Alignment Scores:

Pred. No.:	41.6	Length:	1126
Score:	81.00	Matches:	43
Percent Similarity:	43.48%	Conservative:	27
Best Local Similarity:	26.71%	Mismatches:	58
Query Match:	11.44%	Indels:	34
DB:	23	Gaps:	7

US-10-087-573-2 (1-141) x ABL28661 (1-1126)

QY 8 ThrAsnPhenValAlaGluAsnArgProThrPheGly----- 19

Db 714 TCGACATCGTTGCCCTCGACCTCTGTTGTTGGTGTGCAACGTGCAGAAATCCCAT 655

QY 20 -----GluThrPheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArg 37

Db 654 TGTCTTCAGCATCTCGACCTTCGCGCATTTACCTCCAGGTC-----CGC 607

QY 38 LeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuPro----- 54

Db 606 CTCGTTGATCTTGGCCCGGTC---GCGAGCTGTGCGCCGCGCATTTCTCTCGTCTCT 550

QY 55 -----GlyThrGlyAlaSerAlaAlaAlaThrValThrProLysGlyAla 70

Db 549 TCGCGACGTCGATGTTCCGAGCTGCTGTTGCGCGCTGCTTCGCGCGCGCTCTCTCC 490

QY 71 SerMetLysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeu 90

Db 489 TCCTCTCTCTCCGGGACCCCTCGACCGCATCT---AGATCGCGA--CTCGGCGCAGG 435

QY 91 SerArg-----LysIleArgGluMetAsnLysThrIle 101

Db 434 TCGCGCTCCCGTTCCCGCTCCCGATCTCGGCTGGGACTCGCGTCCGATCCGCTTCCCGG 375

QY 102 SerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLys 121

Db 374 TCGCGGACCGGCGCTCCAGATCCGATCTCGGCTGGGCTACTCCGATGCGCTCC 315

QY 122 ArgAlaGluTyrPheArgHisLeuArgSerLeuLysSer---GlnGlyValAsnArgLeu 140

Db 314 CGCTCCCGGAACGCGCTCCCTCGCTTCTTCGCTCGGATCGCTCTCTCTCGGCTC 255

QY 141 Ile 141

Db 254 CTG 252

RESULT 33

AAH77199

XX ID AAH77199 standard; DNA; 1797 BP.

XX AC AAH77199;

XX DT 24-JAN-2002 (first entry)

XX DE Phaeotrichoconis crotalariae haloperoxidase gene.

XX KW Haloperoxidase; enzyme; halide ion; bactericide; detergent; disinfectant; microbial growth inhibition; preservation agent; enzymatic bleaching; ds.

XX OS Phaeotrichoconis crotalariae.

XX FH Location/Qualifiers

FT CDS 1..1797

FT /tag= a

FT /product= "Haloperoxidase"

FT /partial

FT /note= "No stop codon given"

XX WO200179462-A2.

PN 25-OCT-2001.

PD 10-APR-2001; 2001WO-DK00245.

PF 14-APR-2000; 2000DK-0000625.

PR (NOVO ) NOVOZYMES AS.

PA (MAXY-) MAXYGEN INC.

XX Danielssen S, Schneider P;

XX WPI; 2002-026023/03.

XX F-PSDB; AAG77921.

XX Isolated nucleic acids which encode polypeptides having haloperoxidase activity which can be used in compositions for anti-bacterial activity

XX Claim 1; Page 47-49; 51pp; English.

XX The sequence represents the novel polynucleotide of the invention encoding a polypeptide having haloperoxidase activity, isolated from Phaeotrichoconis crotalariae. The haloperoxidase of the invention works as an enzyme or by oxidation of halide ions, having bactericidal activity. The nucleotide sequence is useful for DNA shuffling (recombination) such that new polynucleotide sequences obtained may encode new polypeptides having haloperoxidase activity with improved properties. The invention relates to methods for killing or inhibiting the growth of microbial cells. The haloperoxidase of the invention may be added to become a component of a detergent composition. The compositions may also comprise one or more other enzymes such as a protease, a lipase,

CC a cutinase or an amylase. The haloperoxidases may be used as preservation  
 CC agents and disinfectant agents. The haloperoxidase also may be used in  
 CC enzymatic bleaching applications.

SQ Sequence 1797 BP; 379 A; 584 C; 473 G; 361 T; 0 other;

Alignment Scores:  
 Pred. No.: 77 Length: 1797  
 Score: 81.00 Matches: 27  
 Percent Similarity: 38.64% Conservative: 7  
 Best Local Similarity: 30.68% Mismatches: 27  
 Query Match: 11.44% Indels: 27  
 DB: 24 Gaps: 3

US-10-087-573-2 (1-141) x AAH77199 (1-1797)

QY 54 ProGlyThrGlyAlaSerAlaIleAlaThrValThrProLysGlyAlaSerMetLys 73  
 DB 1073 CCGGCTACGGATGACGGCGCTCAGACACCGACCCCTTCTGCTCACCTCGCG 1132  
 QY 74 LeuLysPro-----ProArgProGlnSerThrLysSerProGluLeuArgGlu 89  
 DB 1133 CTCCTGCCAACACAAACAGACATCCCTTCAAGCCCCCTTCCCGCTACCGGCG 1192  
 QY 90 LeuSerArgLysIle-ArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValas 109  
 DB 1193 GCCACGGACCTTTGGCGG----- 1211  
 QY 109 nHisArgLeuProGluGlyHisProLeuLeuGluLys----- 121  
 DB 1212 -CGCGTCTTCCAATGGTGGCGGCTACTACACGGCGGCTCGGACGTGGAGGACA 1270  
 QY 122 -ArgAlaGluTyrPheArgHis 128  
 DB 1271 ACGAGCCGACAAACATCGCCAT 1292

RESULT 34

AAH77155  
 ID AAH77155 standard; DNA; 1797 BP.

AC AAH77155;

DT 21-JAN-2002 (first entry)

DE Haloperoxidase coding sequence.

XX Haloperoxidase; antibacterial; microbial growth inhibition; preservation;  
 KW disinfectant; enzymatic bleaching; pulp bleaching; stain bleaching; ds.

OS Phaeotrichoconis crotalariae.

FH Key Location/Qualifiers  
 FT CDS 1..1797

FT /\*tag= a  
 FT /product= "Haloperoxidase"  
 FT /partial  
 FT /note= "No stop codon given"

XX W0200179461-A2.

XX 25-OCT-2001.

XX 10-APR-2001; 2001WO-DK00243.

XX 14-APR-2000; 2000DK-0000625.

PA (NOVO ) NOVOZYMES A/S.

XX Danielson S, Schneider P;

XX WPI; 2002-017611/02.

XX P-PSDB; AAG77905.

XX

PT New haloperoxidase purified from Phaeotrichoconis crotalariae has  
 PT bactericidal activity and is useful as a preservative or disinfectant,  
 PT for example in body care products and food preparation -  
 XX  
 PS Claim 1; Page 46-48; 50pp; English.

CC The sequence represents the coding sequence for the novel polypeptide  
 CC of the invention having haloperoxidase activity. The polypeptide of the  
 CC invention also has antibacterial activity. The haloperoxidase is used to  
 CC kill or inhibit growth of microbial cells. It may be used to preserve or  
 CC disinfect, for example in water based paints or personal care products,  
 CC for cleaning surfaces and cooking utensils in food processing plants and  
 CC any area where food is prepared or served. It may also be used in  
 CC enzymatic bleaching applications, for example pulp or stain bleaching.

XX SQ Sequence 1797 BP; 379 A; 584 C; 473 G; 361 T; 0 other;

Alignment Scores:  
 Pred. No.: 77 Length: 1797  
 Score: 81.00 Matches: 27  
 Percent Similarity: 38.64% Conservative: 7  
 Best Local Similarity: 30.68% Mismatches: 27  
 Query Match: 11.44% Indels: 27  
 DB: 24 Gaps: 3

US-10-087-573-2 (1-141) x AAH77155 (1-1797)

QY 54 ProGlyThrGlyAlaSerAlaIleAlaThrValThrProLysGlyAlaSerMetLys 73  
 DB 1073 CCGGCTACGGATGACGGCGCTCAGACACCGACCCCTTCTGCTCACCTCGCG 1132  
 QY 74 LeuLysPro-----ProArgProGlnSerThrLysSerProGluLeuArgGlu 89  
 DB 1133 CTCCTGCCAACACAAACAGACATCCCTTCAAGCCCCCTTCCCGCTACCGGCG 1192  
 QY 90 LeuSerArgLysIle-ArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValas 109  
 DB 1193 GCCACGGACCTTTGGCGG----- 1211  
 QY 109 nHisArgLeuProGluGlyHisProLeuLeuGluLys----- 121  
 DB 1212 -CGCGTCTTCCAATGGTGGCGGCTACTACACGGCGGCTCGGACGTGGAGGACA 1270  
 QY 122 -ArgAlaGluTyrPheArgHis 128  
 DB 1271 ACGAGCCGACAAACATCGCCAT 1292

RESULT 35

ABK84630/C

ID ABK84630 standard; cDNA; 3530 BP.

XX ABK84630;

XX 14-AUG-2002 (first entry)

DE Human cDNA differentially expressed in granulocytic cells #1201.

XX Human; ss; granulocytic cell; DNA chip; bacterial infection;  
 KW viral infection; parasitic infection; protozoal infection;  
 KW fungal infection; sterile inflammatory disease; psoriasis;  
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
 KW adult respiratory distress syndrome; inflammatory bowel disease;  
 KW Crohn's disease; ulcerative colitis; periodontal disease;  
 KW granulocyte activation; chronic inflammation; allergy.

XX Homo sapiens.

XX WO200228999-A2.

XX 11-APR-2002.

XX 03-OCT-2001; 2001WO-US30821.

03-OCT-2000; 2000US-237189P.  
(GENE-) GENE LOGIC INC.  
Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
WPI; 2002-435328/46.  
Detecting granulocyte activation by detecting differential expression  
of genes associated with granulocyte activation, which serves as  
diagnostic markers that is useful for monitoring disease states and  
drug toxicity -  
Claim 1: SEQ ID No 1201; 114pp; English.  
The invention relates to detecting (M1) granulocyte (GC) activation  
(GCA), by detecting the level of expression of gene(s) (Gs) identified by  
DNA chip analysis as given in the specification, and comparing  
the expression level to an expression level in an unactivated  
GC, where differential expression of Gs is indicative of GCA.  
Also included are modulating (M2) GA by contacting GC with an agent  
that alters the expression of at least one gene in Gs; (2) screening (M3)  
for an agent capable of modulating GCA or an inflammation (especially  
chronic) in a tissue, an allergic response in a subject, exposure of a  
subject to a pathogen or sterile inflammatory disease using the  
gene expression profile; (3) detecting (M4) an inflammation (especially  
chronic) in a tissue, an allergic response in a subject, exposure of a  
subject to a pathogen or sterile inflammatory disease, by detecting the  
level of expression in a sample of the tissue of gene(s) from Gs, where  
(4) treating (M5) an inflammation (especially chronic) or in a tissue,  
an allergic response in a subject, exposure of a subject to a pathogen  
or sterile inflammatory disease, by contacting a tissue having  
inflammation with an agent that modulates the expression of gene(s)  
from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for  
modulating GA; M3 is useful for screening an agent capable of modulating  
GCA preferably in an inflammation in a tissue; M4 is useful for  
detecting an inflammation (especially chronic) in a tissue, an allergic  
response in a subject, exposure of a subject to a pathogen or sterile  
inflammatory disease (e.g. psoriasis, rheumatoid arthritis,  
glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal  
reperfusion injury, ARDS, adult respiratory distress syndrome,  
inflammatory bowel disease, Crohn's disease, ulcerative colitis,  
periodontal disease; also bacterial infection, viral infection,  
parasitic infection, protozoal infection, fungal infection and M5 is  
useful for treating one of the above conditions. The present  
sequence represents a gene differentially expressed in granulocytes.  
Note: The sequence data for this patent did not form part  
of the printed specification, but was obtained in electronic  
format directly from WIPO at  
ftp.wipo.int/pub/published\_pct\_sequences.  
Sequence 3530 BP; 735 A; 902 C; 1044 G; 848 T; 0 other;

```

Alignment Scores:
Pred. No.:      187      Length:      3530
Score:          81.00     Matches:      330
Percent Similarity: 38.04% Conservative:  5
Best Local Similarity: 32.61% Mismatches:   32
Query Match:      11.44% Indels:        25
DB:              24      Gaps:         2

US-10-087-573-2 (1-141) x ABX84630 (1-3530)

Qy      31  ArgValLySerSerGIuArgLeuAlaMet-LeuArgAlaLeuAlaGlyMetCysGlyHi 50
      |||||
Db      595  CGCACGCTCAGTCTCCAGGTGCTGCCACCGCGCACTCGGAGATGACAGCATTA 536

Qy      50  sArgValLeu-----ProGlyThrGlyAlaSerAlaIleAlaThrVa 65
      |||||
Db      535  CCAGATTGTCAATACATACGTCCGCCACAGCGGGGGCTCTCTCTCCGGCAGCCCATGTT 476

```

Qy	65	lthrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSerThrLysSerPr	85
Dd	475	GCGCCCGCTCGAGCGCGCAGCCGCCGGAAACCGGAGCCCGAGGCCTTCGACGCC	416
Qy	85	oGluleuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleserGlnGluSe	105
Dd	415	CCAG-----	412
Qy	105	rLaargValasnHisArgLeuProGluGlyHis	116
Dd	411	-AGACGGGCTGGCGCGCGCGGAACCCGAGGGCCAC	379
 RESULT 36 ABN97276/c ID ABN97276 standard; DNA; 3530 BP.			
XX	ABN97276;		
XX	AC AC		
DT	13-AUG-2002 (first entry)		
XX	XX	Gene #3774 used to diagnose liver cancer.	
DE	XX	Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;	
KW	XX	metastatic liver tumour; cytostatic; expression profile; disease state;	
KW	XX	disease progression; drug toxicity; drug efficacy; drug metabolism.	
OS	XX	Homo sapiens.	
XX	XX	WO200229103-A2.	
PN	XX		
PD	XX	11-APR-2002.	
XX	XX		
PF	XX	02-OCT-2001; 2001WO-US30589.	
XX	XX		
PR	XX	02-OCT-2000; 2000US-237054P.	
XX	XX	(GENE-) GENE LOGIC INC.	
FA	XX		
XX	XX		
Pi	XX	Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;	
DR	XX	WPI; 2002-426119/45.	
XX	XX		
PT	XX	Diagnosing and detecting the progression of liver cancer,	
PT	XX	hepatocellular carcinoma or metastatic liver tumor in a patient,	
PT	XX	involves detecting the level of expression of two or more genes in a	
XX	XX	liver tissue sample -	
PS	XX	Claim 1; SEQ ID NO 3774; 298pp; English.	

The invention relates to a novel method for diagnosing and detecting the progression of liver cancer, hepatocellular carcinoma or metastatic liver tumour in a patient, and differentiating metastatic liver cancer from hepatocellular carcinoma in a patient, involving detecting the level of expression of two or more genes represented in AEN93503-ABN97455 in a tissue sample. The method of the invention has hepatocytic, and/or cytotactic activity. The method is useful for diagnosing and detecting the progression of liver cancer, hepatocellular carcinoma and metastatic liver carcinoma in a patient. The method is useful for identifying expression profiles which serve as useful diagnostic markers as well as markers that can be used to monitor disease states, disease progression, drug toxicity, drug efficacy and drug metabolism.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published/pct\\_sequences](http://ftp.wipo.int/pub/published/pct_sequences).

XX	Sequence	3530 BP; 736 A; 902 C; 1044 G; 848 T; 0 other;
Alignment Scores:		
Pred. No.:	187	Length: 3530
Score:	81.00	Matches: 30
Percent Similarity:	38.04%	Conservative: 5
Best Local Similarity:	32.61%	Mismatches: 32
Query Match:	11.44%	Indels: 25





```

Db      415 CCAG----- 412
QY      105 rAlaArgValaenHisArgLeuProGluGlyHis 116
Db      411 -AGACGGGCTGGCGCGCCGACACCGAGGGCCAC 379

RESULT 38
ABL69756/c
ID ABL69756 standard; DNA; 3530 BP.
XX
AC ABL69756;
XX
XX 15-MAY-2002 (first entry)
XX
DE Prostate cancer related gene sequence SEQ ID NO:8093.
XX
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200194629-A2.
PN
XX
XX 13-DEC-2001.
PD
XX
XX 30-MAY-2001; 2001WO-US10838.
PF
XX
XX 05-JUN-2000; 2000US-209473P.
PR
XX 05-JUN-2000; 2000US-209531P.
PR
XX 18-SEP-2000; 2000US-233133P.
PR
XX 18-SEP-2000; 2000US-233617P.
PR
XX 20-SEP-2000; 2000US-234009P.
PR
XX 20-SEP-2000; 2000US-234034P.
PR
XX 20-SEP-2000; 2000US-234052P.
PR
XX 22-SEP-2000; 2000US-234059P.
PR
XX 22-SEP-2000; 2000US-234567P.
PR
XX 25-SEP-2000; 2000US-234923P.
PR
XX 25-SEP-2000; 2000US-234924P.
PR
XX 25-SEP-2000; 2000US-235077P.
PR
XX 25-SEP-2000; 2000US-235082P.
PR
XX 25-SEP-2000; 2000US-235134P.
PR
XX 25-SEP-2000; 2000US-235280P.
PR
XX 26-SEP-2000; 2000US-235637P.
PR
XX 26-SEP-2000; 2000US-235638P.
PR
XX 27-SEP-2000; 2000US-235711P.
PR
XX 27-SEP-2000; 2000US-235720P.
PR
XX 27-SEP-2000; 2000US-235840P.
PR
XX 27-SEP-2000; 2000US-235863P.
PR
XX 28-SEP-2000; 2000US-236028P.
PR
XX 28-SEP-2000; 2000US-236032P.
PR
XX 28-SEP-2000; 2000US-236033P.
PR
XX 28-SEP-2000; 2000US-236034P.
PR
XX 28-SEP-2000; 2000US-236109P.
PR
XX 28-SEP-2000; 2000US-236111P.
PR
XX 29-SEP-2000; 2000US-236842P.
PR
XX 29-SEP-2000; 2000US-236891P.
PR
XX 02-OCT-2000; 2000US-237172P.
PR
XX 02-OCT-2000; 2000US-237173P.
PR
XX 02-OCT-2000; 2000US-237278P.
PR
XX 02-OCT-2000; 2000US-237294P.
PR
XX 02-OCT-2000; 2000US-237295P.
PR
XX 02-OCT-2000; 2000US-237425P.
PR
XX 03-OCT-2000; 2000US-237598P.
PR
XX 03-OCT-2000; 2000US-237604P.
PR
XX 03-OCT-2000; 2000US-237605P.
PR
XX 03-OCT-2000; 2000US-237608P.
PR
XX 01-NOV-2000; 2000US-244867P.
PR
XX 01-NOV-2000; 2000US-245084P.
XX

```

```

PA (AVAL-) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
XX MPI; 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set
XX
XX Claim 1; SEQ ID 8093; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
XX anti-neoplastic agent. The method involves exposing cells to a chemical
XX agent to be tested for anti-neoplastic activity, determining a change in
XX expression of at least one gene (I) of a signature gene set, where (I)
XX comprises a sequence (S) selected from 8447 sequences (given in ABL61664
XX to ABL70110), or is at least 95% identical to (S), where a change in
XX expression is indicative of anti-neoplastic activity. (I) has cytostatic
XX activity and can be used in gene therapy. M1 can be used for screening
XX an anti-neoplastic agent, and can be used for producing a product which
XX is the data collected with respect to the anti-neoplastic agent as a
XX result of M1, and the data is sufficient to convey the chemical
XX structure and/or properties of the agent. M1 can be used in the
XX treatment of cancer such as colon, breast, stomach, lung, thyroid,
XX oesophageal, ovarian, kidney, prostate or pancreatic cancer,
XX adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
XX infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
XX carcinoma, papillary carcinoma and Wilm's tumour.
XX
XX Sequence 3530 BP; 736 A; 902 C; 1044 G; 848 T; 0 other;
SQ

Alignment Scores:
Pred. No.: 187 Length: 3530
Score: 81.00 Matches: 30
Percent Similarity: 38.04% Conservative: 5
Best Local Similarity: 32.61% Mismatches: 32
Query Match: 11.44% Indels: 25
DB: 24 Gaps: 2

US-10-087-573-2 (1-141) x ABL69756 (1-3530)
QY 31 ArgValLysSerSerGluArgLeuAlaMet-LeuArgAlaLeuAlaGlyMetCysGlyHi 50
Db 595 CGCACGGCTCAGCTTCACAGGTGGCTTCCCGCCGCGGCTCTCTCGCGGAGCCATGTT 536
QY 50 sArgValLeu-----ProGlyThrGlyAlaSerAlaAlaLeuAlaThrVa 65
Db 535 CCGAGTTGTCATCATACACGTCGCGCCGCGGCGGCTCTCTCGCGGAGCCATGTT 476
QY 65 lThrProLysGlyAlaSerMetLysLeuLysProArgProGlnSerThrLysSerPr 85
Db 475 GGCGCCCGTCGAGCGCGGAGCCGCGCGCCGCGGAGCCGAGAGCCGAGAGCTTCGAGCCC 416
QY 85 oGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSe 105
Db 415 CCAG----- 412
QY 105 rAlaArgValAsnHisArgLeuProGluGlyHis 116
Db 411 -AGACGGGCTGGCGCGCCGACACCGAGGGCCAC 379

RESULT 39
ABL26682/c
ID ABL26682 standard; DNA; 6105 BP.
XX
XX ABL26682;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 31519.
DE
XX
XX

```

KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
OS Drosophila melanogaster.  
XX WO200171042-A2.  
PN 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US09231.  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX (PEKE ) PE CORP NY.  
PA Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI; 2001-656860/75.  
DR New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX Claim 1; SEQ ID NO 31519; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (BB57737-AB572072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 6105 BP; 1624 A; 1403 C; 1377 G; 1701 T; 0 other;

Alignment Scores:  
Pred. No.: 385 Length: 6105  
Score: 81.00 Matches: 33  
Percent Similarity: 39.32% Conservative: 13  
Best Local Similarity: 28.21% Mismatches: 31  
Query Match: 11.44% Indels: 40  
DB: 23 Gaps: 4

US-10-087-573-2 (1-141) x ABL26682 (1-6105)  
Qy 20 GluThrPheAspValMetArg-----GluAlaLeuLeuArg 31  
Db 4274 GAAGATTTCAGCGCCATGAGAGGAGGATCTGAAATGCGCATCGAGAGATGGTTAG 4215  
Qy 32 Vallys---SerSerGluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHis 50  
Db 4214 ATCAGGTGTAGTACCGAGAGACTCTTATAATAGT----- 4179  
Qy 51 ArgValLeuProGlyThrGlyAlaSerAlaAlaLeuValThrProLysGlyAla 70  
Db 4179 ----- 4179  
Qy 71 SerMetLysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeu 90  
Db 4178 AGCCAGAGAGCTCAAACTCTTCATCATCAGAGCACCGTGTCCGTGGATTCGCGGATTTG 4119  
Qy 91 SerArgLysLeuArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHis 110  
Db 4118 GAGTCCGCTGCCAGAGAGCTGCAATTCGACATTAGCCAGTACACACAGAGATCGAGGAG 4059  
Qy 111 ArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGluTyrPheArg 127  
Db 4058 CTCAAACAGGAA-----CTGCTTAGGAGGCAACACAGAACTGGAGCC 4017

RESULT 40  
ABQ51890/c  
ID ABQ51890 standard; DNA; 681 BP.  
XX ABQ51890;  
AC ABQ51890;  
XX 12-JUL-2002 (first entry)  
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 38481.  
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX Homo sapiens.  
OS WO200218632-A2.  
XX 07-MAR-2002.  
PD 01-SEP-2001; 2001WO-EP10074.  
XX 01-SEP-2000; 2000DE-1043826.  
PR 05-SEP-2000; 2000DE-1044543.  
XX (EPIG-) EPIGENOMICS AG.  
PA Olek A, Piepenbrock C, Berlin K, Guetig D;  
PI WPI; 2002-371829/40.  
XX Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA -  
XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ34410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.  
XX SQ Sequence 681 BP; 67 A; 83 C; 251 G; 280 T; 0 other;  
Alignment Scores:  
Pred. No.: 24.3 Length: 681  
Score: 80.50 Matches: 23  
Percent Similarity: 41.03% Conservative: 9  
Best Local Similarity: 25.49% Mismatches: 35  
Query Match: 11.37% Indels: 11  
DB: 24 Gaps: 2  
US-10-087-573-2 (1-141) x ABQ51890 (1-681)  
Qy 67 ProLysGlyAlaSerMetLysLeuLysPro-----ProArgPro 79



OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 17, 2003, 12:56:54 : Search time 56 Seconds

(without alignments)  
1111.340 Million cell updates/sec

Title: US-10-087-573-2

Perfect score: 708

Sequence: 1 MESTSTTTFVAENRPTFGE.....RAEYFRHLRLSLKSGVNRLLI 141

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US1008753/tunat\_14112003\_103611\_20707/app\_query.fasta\_1.327  
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPECL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bblosum62 -TRANS=human40  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=pco -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US1008753@cgn\_1\_1\_44 @tunat\_14112003\_103611\_20707 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	13.0	774	US-09-252-991A-15297	Sequence 15297, A
2	92	13.0	2163	US-09-252-991A-15344	Sequence 15344, A
3	92	13.0	4260	US-09-252-991A-15251	Sequence 15251, A
C 4	84.5	11.9	1530	US-09-252-991A-6770	Sequence 6770, Ap
5	84.5	11.9	1695	US-09-252-991A-6969	Sequence 6969, Ap
C 6	83	11.7	1677	US-09-252-991A-4734	Sequence 4734, Ap
7	83	11.7	2151	US-09-252-991A-4774	Sequence 4774, Ap
C 8	82.5	11.7	750	US-09-252-991A-12401	Sequence 12401, A
9	82.5	11.7	1581	US-09-252-991A-12190	Sequence 12190, A
10	82.5	11.7	1620	US-09-252-991A-12255	Sequence 12255, A
11	81	11.4	1797	US-09-833-441-1	Sequence 1, Appli
12	81	11.4	1797	US-09-833-102-1	Sequence 1, Appli

13	80	11.3	1422	4	US-09-252-991A-14835	Sequence 14835, A
14	80	11.3	1446	4	US-09-252-991A-15091	Sequence 15091, A
C 15	79.5	11.2	1149	3	US-08-660-645A-8	Sequence 8, Appli
C 16	79.5	11.2	1149	3	US-09-298-718-8	Sequence 8, Appli
C 17	79.5	11.2	1149	3	US-09-546-969-8	Sequence 8, Appli
18	79.5	11.2	8625	3	US-08-980-832-1	Sequence 1, Appli
C 19	79.5	11.2	11233	3	US-08-980-832-27	Sequence 27, Appli
C 20	78.5	11.1	997	4	US-09-452-239-13	Sequence 13, Appli
21	78.5	11.1	4403765	3	US-09-103-840A-2	Sequence 2, Appli
22	78.5	11.1	441529	3	US-09-103-840A-1	Sequence 1, Appli
23	78	11.0	2061	4	US-09-107-532A-1348	Sequence 1348, Ap
24	77.5	10.9	3856	2	US-07-743-357-20	Sequence 20, Appli
25	77.5	10.9	4113	2	US-07-743-357-21	Sequence 21, Appli
26	77.5	10.9	8932	3	US-09-124-900-1	Sequence 1, Appli
27	77.5	10.9	8933	3	US-08-463-210-4	Sequence 4, Appli
28	77.5	10.9	8933	3	US-09-620-958A-3	Sequence 3, Appli
29	77.5	10.9	8933	3	US-09-620-958A-4	Sequence 4, Appli
30	77.5	10.9	8933	3	US-09-620-958A-9	Sequence 9, Appli
31	77.5	10.9	9719	4	US-09-700-304-1	Sequence 1, Appli
32	77.5	10.9	11672	4	US-09-441-340-2	Sequence 2, Appli
33	77.5	10.9	15611	4	US-09-441-340-1	Sequence 1, Appli
34	77	10.9	2505	1	US-07-977-434-7	Sequence 7, Appli
35	77	10.9	2505	1	US-08-458-819-7	Sequence 7, Appli
C 36	77	10.9	2505	5	PCT-US91-07035-7	Sequence 7, Appli
C 37	77	10.9	68750	3	US-09-335-409-1	Sequence 1, Appli
C 38	77	10.9	68750	4	US-09-568-102-1	Sequence 1, Appli
C 39	77	10.9	68750	4	US-09-567-969-1	Sequence 1, Appli
C 40	77	10.9	68750	4	US-09-568-480-1	Sequence 1, Appli
C 41	77	10.9	68750	4	US-09-568-486-1	Sequence 1, Appli
C 42	77	10.9	68750	4	US-09-568-472-1	Sequence 1, Appli
C 43	77	10.9	68750	4	US-09-567-899-1	Sequence 1, Appli
C 44	77	10.9	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C 45	77	10.9	4411529	3	US-09-103-840A-1	Sequence 1, Appli

#### ALIGNMENTS

##### RESULT 1

US-09-252-991A-15297  
; Sequence 15297, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15297  
; LENGTH: 774  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15297

Alignment Scores:  
Pred. No.: 0.192 Length: 774  
Score: 92.00 Matches: 44  
Percent Similarity: 46.03% Conservative: 14  
Best Local Similarity: 34.92% Mismatches: 51  
Query Match: 12.99% Indels: 17  
DB: 4 Gaps: 5

US-10-087-573-2 (1-141) x US-09-252-991A-15297 (1-774)

QY 25 MetArgGluAlaLeuArgVallySerSerGluArg-----LeuAla----- 39

DB 321 CTGCGGAGAGAGTTGCTGCGCGCCGCGAATGTCGCGCCAGCTCGCGACACCGCGGAA 380



; Sequence 6770, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6770  
; LENGTH: 1530  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6770

Alignment Scores:  
Pred. No.: 3.5 Length: 1530  
Score: 84.50 Matches: 43  
Percent Similarity: 41.40% Conservative: 22  
Best Local Similarity: 27.39% Mismatches: 71  
Query Match: 11.94% Indels: 22  
Gaps: 4

US-10-087-573-2 (1-141) x US-09-252-991A-6770 (1-1530)

QY 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgPro---ThrPheGly 19  
Db 1130 ATGAACCGGTGAGCCAGACG-----GCGCGGGTGGCGCGATGATTCGTT 1083  
QY 20 GluThrPhe-----AspValMetArgGluAlaLeuArgValLysSerSerGluArg 37  
Db 1082 GCGATGTTCTGGAAGTGGTTCGGCGGAGACGATGTTCCGCCGCTGTTGTCGAGAG 1023  
QY 38 LeuAla---MetLeuArgAlaLeuAla-----GlyMetCys--- 48  
Db 1022 GTGGCAGCATGATCAACGCCATCGCCACCGCCAGCCCGGGGTGGGATGTCGCCG 963  
QY 49 -----GlyHisArgValLeuProGlyThrGlyAlaSerAlaLeuAlaThr 64  
Db 962 ATCAGGCTGACCGGGTGGAAACCGTCTCGCCACCCAGCGAAGCCGAGCGCGGTCTATC 903  
QY 65 ValThrProLysGlyAlaSerMetLysLeuLysProArgProGlnSerThrLysSer 84  
Db 902 ACCACCGCGAGCGCGGACAGACATGTCAGCGGAGCGCGGATCTGCCCGAGT 843  
QY 85 ProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGlu 104  
Db 842 AT-CTGGTCTCTTGGCTACGCGCATAGCGCTGAAAGTCGGGATGTTTCAGCGACAGCGT 784  
QY 105 SerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGlu 124  
Db 783 GGCCCGAAGCCAGCATCCCTCAGCCCGCCATGAAGTACCCGGTGAC---GCTGGC 727  
QY 125 TyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeuLeu 141  
Db 726 GCCTTCGGCGGCTTCGGCGGCTGCGCGAGCAGTTCGCTGACCGATATGTG 676

RESULT 5  
US-09-252-991A-6969  
; Sequence 6969, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6969  
; LENGTH: 1695  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6969

Alignment Scores:  
Pred. No.: 4.04 Length: 1695  
Score: 84.50 Matches: 43  
Percent Similarity: 41.40% Conservative: 22  
Best Local Similarity: 27.39% Mismatches: 71  
Query Match: 11.94% Indels: 22  
Gaps: 4

US-10-087-573-2 (1-141) x US-09-252-991A-6969 (1-1695)

QY 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgPro---ThrPheGly 19  
Db 434 ATGAACCGGTGAGCCAGACG-----GCGCGGGTGGCGCGATGATTCGTT 481  
QY 20 GluThrPhe-----AspValMetArgGluAlaLeuArgValLysSerSerGluArg 37  
Db 482 GCGATGTTCTGGAAGTGGTTCGGCGGAGACGATGTTCCGCCGCTGTTGTCGAGAGG 541  
QY 38 LeuAla---MetLeuArgAlaLeuAla-----GlyMetCys--- 48  
Db 542 GTGGCAGCATGATCAACGCCATCGCCACCCAGCCCGGGGTGGGATGTCGCCG 601  
QY 49 -----GlyHisArgValLeuProGlyThrGlyAlaSerAlaLeuAlaThr 64  
Db 602 ATCAGGCTGACCGGGTGGAAACCGTCTCGCCACCCAGCGAAGCCGAGCGCGGTCTATC 661  
QY 65 ValThrProLysGlyAlaSerMetLysLeuLysProArgProGlnSerThrLysSer 84  
Db 662 ACCACCGCGAGCGCGGACAGACATGTCAGCGGAGCGCGGATCTGCCCGAGT 721  
QY 85 ProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGlu 104  
Db 722 AT-CTGGTCTCTTGGCTACGCGCATAGCGCTGAAAGTCGGGATGTTTCAGCGACAGCGT 780  
QY 105 SerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGlu 124  
Db 781 GGCCCGAAGCGCGACCATCGCGCTCAGCCCGCCATGAAGTAGCCGGTGAC---GCTGGC 837  
QY 125 TyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeuLeu 141  
Db 838 GCCTTCGGCGGCTTCGGCGGCTGCGCGAGCAGTTCGCTGACCGATATGTG 888

RESULT 6  
US-09-252-991A-4734/C  
; Sequence 4734, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 4734  
; LENGTH: 1677  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-4734

Alignment Scores:

Pred. No.: 5.87 Length: 1677  
Score: 83.00 Matches: 35  
Percent Similarity: 35.17% Conservative: 16  
Best Local Similarity: 24.14% Mismatches: 42  
Query Match: 11.72% Indels: 52  
DB: 4 Gaps: 6

US-10-087-573-2 (1-141) x US-09-252-991A-4734 (1-1677)

```
QY 35 SerGluArgLeuAlaMetLeu-----ArgAlaLeuAlaGlyMetCysGly 49
Db 1345 TCCGGCGGTCTGCGGTTCCTCGTTCAAGAACAAAGGGCGTTCCGCTGTTCTCGATGCCGT 1286
QY 50 HisArgValLeuProGlyThr-----56
Db 1285 CATCGACTACCTCCGGCTCCGACCGAGATCCCTGCGATCAAGGGCTGACGGCGGACGA 1226
QY 57 -----GlyAlaSerAlaAlaAlaThrValThrProLysGlyAla 70
Db 1225 CGAGACCGTCGAAGACGAGCGTCTGCGCGACGACACGACCGTTCTCGTCCCTGGCGTT 1166
QY 71 SerMetLysLeuLysProPro-----ArgProGlnSerThrLysSer 84
Db 1165 CAAGATCGCTACCGACCGCTGCTCGGTACCTGACCTTCCCGCGCTTTACTCCGCGT 1106
QY 85 ProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIle-----101
Db 1105 CCTGAGTTCGCGGACTCCGTCCTGAACCTCGTGAAGAGCAAGAGCGCGTCGTCG 1046
QY 102 -----SerGlnGluSerAlaArgValAsnHis 110
Db 1045 GATGTGCAGATGCACGCCAACCGAGCTGAAGAGATCAAGGAAGTGCAGCGGACAT 986
QY 111 ArgLeuPro-----GluGly-----HisProLeuLeuGluLysArg 122
Db 985 CGTCCCTGATCGGCATGAAGGACGTCAACCGCGTGACACCGCTGTGCTCGATCGAGAA 926
QY 123 AlaGluTyrPheArg 127
Db 925 GCCGATCATCTCGA 911
```

RESULT 7

US-09-252-991A-4774  
; Sequence 4774, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 4774  
; LENGTH: 2151  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa

Alignment Scores:

Pred. No.: 8.35 Length: 2151  
Score: 83.00 Matches: 35  
Percent Similarity: 35.17% Conservative: 16  
Best Local Similarity: 24.14% Mismatches: 42  
Query Match: 11.72% Indels: 52  
DB: 4 Gaps: 6

US-10-087-573-2 (1-141) x US-09-252-991A-4774 (1-2151)

```
QY 35 SerGluArgLeuAlaMetLeu-----ArgAlaLeuAlaGlyMetCysGly 49
Db 816 TCCGGCGGTCTGCGGTTCCTCGTTCAAGAACAAAGGGCGTTCCGCTGTTCTCGATGCCGT 875
QY 50 HisArgValLeuProGlyThr-----56
Db 876 CATCGACTACCTCCCGGCTCCGACCGAGATCCCTGCGATCAAGGGCGTCAGCCCGGACGA 935
QY 57 -----GlyAlaSerAlaAlaAlaThrValThrProLysGlyAla 70
Db 936 CGAGACCGTCGAAGACGAGCGTCTGCGCGACGACGACAAACCGTTCTCGTCCCTGGCGTT 995
QY 71 SerMetLysLeuLysProPro-----ArgProGlnSerThrLysSer 84
Db 996 CAAGATCGCTACCGACCGCTTCTGCTGTTACCTGACCTTCCGCCCGCTTTACTCCGCGT 1055
QY 85 ProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIle-----101
Db 1056 CCTGAGTTCGCGGACTCCGTCCTGAACCTCGTGAAGAGCAAGAGCGCGTCGTCG 1115
QY 102 -----SerGlnGluSerAlaArgValAsnHis 110
Db 1116 GATGTGCAGATGCACGCCAACCGAGCTGAAGAGATCAAGGAAGTGCAGCGGACAT 1175
QY 111 ArgLeuPro-----GluGly-----HisProLeuLeuGluLysArg 122
Db 1176 CGCTGCCCTGATCGGCATGAAGGACGTCAACCGCGTGACACCGCTGTGCTCGATCGAGAA 1235
QY 123 AlaGluTyrPheArg 127
Db 1236 GCCGATCATCTCGA 1250
```

RESULT 8

US-09-252-991A-12401/c  
; Sequence 12401, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 12401  
; LENGTH: 750  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa

Alignment Scores:

Pred. No.: 2.14 Length: 750  
Score: 82.50 Matches: 36  
Percent Similarity: 39.71% Conservative: 18  
Best Local Similarity: 26.47% Mismatches: 60  
Query Match: 11.65% Indels: 22  
DB: 4 Gaps: 4

US-10-087-573-2 (1-141) x US-09-252-991A-12401 (1-750)

```
QY 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42
Db 468 GATGGCCCGGAGAGACGACATGTGTACGCCCGCGTCCCGCGCCCGCCGAGACGCCGCCGC 409
QY 43 AlaLeuAlaGlyMetCysGlyHisArgVal-----LeuProGlyThrGlyAlaSerAla 60
Db: 4
```

Db 408 GGAGCCTCCGCGAAGCAGGAGCAGAGCGGCGCGCTGGCGCAACTGGTCAAGCA 349  
Qy 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysPro----- 76  
Db 348 GCAGCGACAACAGTCTCGCGCCCTGGCGCGCAGCAGGAGCAGCGCTGGCGCGCTCGC 289  
Qy 77 -----ProArgProGlnSerThrLysSer 84  
Db 288 CCGCAGCAGCAGAGAGAACTGGCAGCGCGCGAACAGGACGCCCGCGGCAACTGGACAT 229  
Qy 85 ProGluLeuArgGlu-LeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnG 104  
Db 228 CCGCAGCAGGAACTGCTCAGC-----CTGCAACAGGCCCTGGAAACGGCAGCGCGGGA 175  
Qy 104 userAlaArgValAsnHisArgLeuProGlu---GlyHisProLeuLeuGluLysArgAl 123  
Db 174 GAACGCCGAATTCAGCAGCGCTCTGCTGGAGCAGCGGCGAGCAGTTCAGCGCAACCGCGA 115  
Qy 123 aGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsn 138  
Db 114 GGAACCTACCCGCAACTGCGTTTCATCGAGAACCGGCGCGCAAC 69  
RESULT 9  
US-09-252-991A-12190  
; Sequence 12190, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 12190  
; LENGTH: 1581  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-12190  
Alignment Scores:  
Pred. No.: 6.15 Length: 1581  
Score: 82.50 Matches: 36  
Percent Similarity: 39.71% Conservative: 18  
Best Local Similarity: 26.47% Mismatches: 60  
Query Match: 11.65% Indels: 22  
DB: 4 Gaps: 4  
US-10-087-573-2 (1-141) x US-09-252-991A-12190 (1-1581)  
Qy 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
Db 541 GATGGCCCGCGAAGAGCAATGTAGCGCGCGCTCCGCGCGCCCGCGAGGACCGCGCGC 600  
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgVal-----LeuProGlyThrGlyAlaSerAla 60  
Db 601 GGAGCCTCCGCGCAAGCAGCAGGAGCAGCAGCAGCGCGCGCTGGCGCAACTGGTCAAGCA 660  
Qy 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysPro----- 76  
Db 661 GCAGCGACAACAGTCTCGCGCGCGCTGGCGCGCAGCAGCAGGACGCGCTGGCGCGGCTCGC 720  
Qy 77 -----ProArgProGlnSerThrLysSer 84  
Db 721 CCGGCGAGCAGCAAGAGAACTGGCGCGCGCGAACAGCAGGACGCCCGCGGCAACTGGACAT 780  
Qy 85 ProGluLeuArgGlu-LeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnG 104  
Db 781 CCGCGCAGCGCAAGTGTCTCAGC-----CTGCAACAGGCCCTGGAAACCGGCGCGCGGA 834

Qy 104 userAlaArgValAsnHisArgLeuProGlu---GlyHisProLeuLeuGluLysArgAl 123  
Db 835 GAACGCCGAATTCGACAGCAGCGTCTCTGAGCAGCGCGAGCAGTTCACGCGCAACCGCGA 894  
Qy 123 aGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsn 138  
Db 895 GGAACCTACCCGCAACTGCGTTTCATCGAGAACCGGCGCGCAAC 940  
RESULT 10  
US-09-252-991A-12255  
; Sequence 12255, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 12255  
; LENGTH: 1620  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-12255  
Alignment Scores:  
Pred. No.: 6.36 Length: 1620  
Score: 82.50 Matches: 36  
Percent Similarity: 39.71% Conservative: 18  
Best Local Similarity: 26.47% Mismatches: 60  
Query Match: 11.65% Indels: 22  
DB: 4 Gaps: 4  
US-10-087-573-2 (1-141) x US-09-252-991A-12255 (1-1620)  
Qy 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
Db 699 GATGGCCCGCGAAGAGCAATGTAGCGCGCGCTCCGCGCGCCCGCGAGGACCGCGCGC 758  
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgVal-----LeuProGlyThrGlyAlaSerAla 60  
Db 759 GGAGCCTCCGCGCAAGCAGCAGGAGCAGCAGCAGCGCGCGCTGGCGCAACTGGTCAAGCA 818  
Qy 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysPro----- 76  
Db 819 GCAGCGACAACAGTCTCGCGCGCGCTGGCGCGCAGCAGGAGCAGCGCTGGCGCGGCTCGC 878  
Qy 77 -----ProArgProGlnSerThrLysSer 84  
Db 879 CCGCAGCAGCAGAGAGAACTGGCGCGCGCGAACAGGACGCCCGCGGCAACTGGACAT 938  
Qy 85 ProGluLeuArgGlu-LeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnG 104  
Db 939 CCGCGCAGCGAAGTGTCTCAGC-----CTGCAACAGCGCGCTGGAAACCGGCGCGGGA 992  
Qy 104 userAlaArgValAsnHisArgLeuProGlu---GlyHisProLeuLeuGluLysArgAl 123  
Db 993 GAACGCCGAATTCAGCAGCGCTCTGCTGGAGCAGCGCGAGCAGTTCACGCGCAACCGCGA 1052  
Qy 123 aGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsn 138  
Db 1053 GGAACCTACCCGCAACTGCGTTTCATCGAGAACCGGCGCGCAAC 1098  
RESULT 11  
US-09-832-441-1  
; Sequence 1, Application US/09832441  
; Patent No. 6506585



```

; GENERAL INFORMATION:
; APPLICANT: Daniels, Steffen
; APPLICANT: Schneider, Palie
; TITLE OF INVENTION: Polypeptides having haloperoxidase activity
; FILE REFERENCE: 10019-200-US
; CURRENT APPLICATION NUMBER: US/09/832,441
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 1
; LENGTH: 1797
; TYPE: DNA
; ORGANISM: Phaeotrichoconis crotalariae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1797)
US-09-832-441-1

Alignment Scores:
Pred. No.: 10.9 Length: 1797
Score: 81.00 Matches: 27
Percent Similarity: 38.64% Conservative: 7
Best Local Similarity: 30.68% Mismatches: 27
Query Match: 11.44% Indels: 27
DB: Gaps: 3

US-10-087-573-2 (1-141) x US-09-832-441-1 (1-1797)

Qy 54 ProGlyThrGlyAlaSerAlaIleAlaThrValThrProLysGlyAlaSerMetLys 73
Db 1073 CCGGCGTACGGGATGACGGCGGTCCAGACCGGCGGCTTCTGGCTACCCCTCGGCG 1132
Qy 74 LeuLysPro-----ProArgProGlnSerThrLysSerProGluLeuArgGlu 89
Db 1133 CTCTGCCACAAACACAAACGACATCCCTTCAAGCCCCCTTCCCGCGCTACCCGTCG 1192
Qy 90 LeuSerArgLysIle-ArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValas 109
Db 1193 GCACGCGGACCTTTGGCG----- 1211
Qy 109 nHisArgLeuProGluGlyHisProLeuLeuGluLys----- 121
Db 1212 -CGCCGCTTCCAAATGGTGGCGGTACTACTACACGGGCGGTGGCAGTGGAGGACA 1270
Qy 122 -ArgAlaGluTyrPheArgHis 128
Db 1271 ACGAGCCGACACATCGCCAT 1292

RESULT 13
US-09-252-991A-14835
; Sequence 14835, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14835
; LENGTH: 1422
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14835

Alignment Scores:
Pred. No.: 10.1 Length: 1422
Score: 80.00 Matches: 36
Percent Similarity: 40.46% Conservative: 17
Best Local Similarity: 27.48% Mismatches: 46
Query Match: 11.30% Indels: 32
DB: Gaps: 7

US-10-087-573-2 (1-141) x US-09-252-991A-14835 (1-1422)

Qy 20 GluThrPheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAla 39
Db 859 GAGGACTATCCGTGTGCGCGCGCTTCTCTACCTGAAGCCCGCGGAGAAAGACCCC 918
Qy 40 MetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSer 59
Db 919 TGGGCCCGAGGCGCTGTA-----CGTTCCGCCAGGCGCGCGCGGCGGCGAG 963
Qy 60 AlaIleAlaThrValThrProLysGly---AlaSerMetLysLeuLys----- 75
Db 964 GCGATC-----GTCACCGCGAGCGGCTTCGTCGAGAGAGATCCAGCGGCGTCCAG 1014
Qy 76 ---ProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94
Db 76 ---ProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94

; GENERAL INFORMATION:
; APPLICANT: Daniels, Steffen
; APPLICANT: Schneider, Palie
; TITLE OF INVENTION: Polypeptides having haloperoxidase activity
; FILE REFERENCE: 10019-200-US
; CURRENT APPLICATION NUMBER: US/09/832,441
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 1
; LENGTH: 1797
; TYPE: DNA
; ORGANISM: Phaeotrichoconis crotalariae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1797)
US-09-832-441-1

Alignment Scores:
Pred. No.: 10.9 Length: 1797
Score: 81.00 Matches: 27
Percent Similarity: 38.64% Conservative: 7
Best Local Similarity: 30.68% Mismatches: 27
Query Match: 11.44% Indels: 27
DB: Gaps: 3

US-10-087-573-2 (1-141) x US-09-832-441-1 (1-1797)

Qy 54 ProGlyThrGlyAlaSerAlaIleAlaThrValThrProLysGlyAlaSerMetLys 73
Db 1073 CCGGCGTACGGGATGACGGCGGTCCAGACCGGCGGCTTCTGGCTACCCCTCGGCG 1132
Qy 74 LeuLysPro-----ProArgProGlnSerThrLysSerProGluLeuArgGlu 89
Db 1133 CTCTGCCACAAACACAAACGACATCCCTTCAAGCCCCCTTCCCGCGCTACCCGTCG 1192
Qy 90 LeuSerArgLysIle-ArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValas 109
Db 1193 GCACGCGGACCTTTGGCG----- 1211
Qy 109 nHisArgLeuProGluGlyHisProLeuLeuGluLys----- 121
Db 1212 -CGCCGCTTCCAAATGGTGGCGGTACTACTACACGGGCGGTGGCAGTGGAGGACA 1270
Qy 122 -ArgAlaGluTyrPheArgHis 128
Db 1271 ACGAGCCGACACATCGCCAT 1292

RESULT 12
US-09-833-102-1
; Sequence 1, Application US/09833102
; Patent No. 6521434
; GENERAL INFORMATION:
; APPLICANT: Daniels, Steffen
; APPLICANT: Schneider, Palie
; TITLE OF INVENTION: Nucleic acids encoding polypeptides having haloperoxidase activity
; FILE REFERENCE: 10172-200-US
; CURRENT APPLICATION NUMBER: US/09/833,102
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 1
; LENGTH: 1797
; TYPE: DNA
; ORGANISM: Phaeotrichoconis crotalariae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1797)
US-09-833-102-1

Alignment Scores:
Pred. No.: 10.9 Length: 1797
Score: 81.00 Matches: 27
```

Db 1015 ATCGCACCGCGCCGAG-----ATGCCGCGGATACCGCAAGCTCGCGGAACAGCGG 1068

QY 95 ArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGlu 114

Db 1069 CGCGGCTGAGC-----GTGAACCTTCGCTCCAGGA 1101

QY 115 GlyHisProLeuLeuGluLysArgAla-----GluTyrPhe 126

Db 1102 GGCAGCGCGCTGCTGCACCAACAGGACCTGCTGCAGCTACAGCGTTTGTGACTACCTG 1161

QY 127 ArgHisLeuArgSerLeuLysSerGlnGlyVal 137

Db 1162 CGGCAAAATCGGAAGTTGCAGGAGCGCACCGTG 1194

RESULT 14

US-09-252-991A-15091

; Sequence 15091, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 15091

; LENGTH: 1446

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-15091

Alignment Scores:

Pred. No.: 10.3 Length: 1446

Score: 80.00 Matches: 36

Percent Similarity: 40.46% Conservative: 17

Best Local Similarity: 27.48% Mismatches: 46

Query Match: 11.30% Indels: 32

DB: 4 Gaps: 7

US-10-087-573-2 (1-141) x US-09-252-991A-15091 (1-1446)

QY 20 GluThrPheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAla 39

Db 818 GAGGACTATCCGCTGTCGCGCGCCTGTTCTTACTACCGCGGCGGAGAGAACCCC 877

QY 40 MetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSer 59

Db 878 TGGGCGCCAGCGCCCTGTA-----CGTTCCCGCCAGCGCGCGCGCCAG 922

QY 60 AlaIleAlaAlaThrValThrProLysGly---AlaSerMetLysLeuLys----- 75

Db 923 GGATC-----GTCACCGCAGCGCTGCTGCGCAGAAATCCAGCGCGGTGCGAG 973

QY 76 ---ProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94

Db 974 ATCGCACCGCGCCGCGAG-----ATGCCCGCGAATACCGCAAGCTCGCGGAACAGCGG 1027

QY 95 ArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGlu 114

Db 1028 CGCGGCTGAGC-----GTGAACCTTCGCTCCAGGA 1060

QY 115 GlyHisProLeuLeuGluLysArgAla-----GluTyrPhe 126

Db 1061 GGCAGCGCGCTGCTGCACCAACAGGACCTGCTGCAGTACACCGTTTGTGACTACCTG 1120

QY 127 ArgHisLeuArgSerLeuLysSerGlnGlyVal 137

Db 1121 CGGCAAAATCGGAAGTTGCAGGAGCGCACCGTG 1153

RESULT 15

US-08-660-645A-8/C

; Sequence 8, Application US/08660645A

; Patent No. 6087152

; GENERAL INFORMATION:

; APPLICANT: Hohmann, Hans-Peter

; APPLICANT: Pasamontes, Luis

; APPLICANT: Tessier, Michel

; APPLICANT: van Loon, Adolphus

; TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION

; NUMBER OF SEQUENCES: 47

; CORRESPONDENCE ADDRESS:

; STREET: 340 Kingsland Street

; CITY: Nutley

; STATE: NJ

; COUNTRY: USA

; ZIP: 07110

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA: US/08/660,645A

; APPLICATION NUMBER: US/08/660,645A

; FILING DATE: 07-JUN-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP 95108888.9

; FILING DATE: 09-JUN-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Pokras, Bruce A.

; REGISTRATION NUMBER: 32,748

; REFERENCE/DOCKET NUMBER: RAN 6002/170

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (201) 235-5801

; TELEFAX: (201) 235-2363

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1149 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-08-660-645A-8

Alignment Scores:

Pred. No.: 8.51 Length: 1149

Score: 79.50 Matches: 35

Percent Similarity: 42.15% Conservative: 16

Best Local Similarity: 28.93% Mismatches: 50

Query Match: 11.23% Indels: 20

DB: 3 Gaps: 4

US-10-087-573-2 (1-141) x US-08-660-645A-8 (1-1149)

QY 37 ArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArg----- 51

Db 590 CGCGTATCCTCGATCAGATGCGGCTGGAGTGAAGGCGAGCAGATAGATGACCGGTAC 531

QY 52 ---ValLeuProGlyThrGlyAlaSerAlaIleAlaAlaThrValThrProLysGlyAla 70

Db 530 CCGTCCATCTCGCGAAGCGTCCGCTCCATGATCATCGGGGCGCTCGAGCCATGGGGCGG 471

QY 71 SerMetLysLeu-LysProArg-----ProGlnSerThrLysSerProGln 86

Db 470 TCGGTCTCGATCTCGACCGCCACGAAATTTGGAAACCCAGCGTCAGGTGCGGGGTCTCG 411

QY 86 uLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAl 106

Db 410 A---CGGCACACCGGGGCTCGATCAGCAGCGACCGCTCGATCCGCGAGCGCTCGGTGAGC 354

Qy 106 aArgValAsnHisArg-----LeuProGluGlyHisProLe 118  
 Db 353 GTCCGCGCGGTATGTCACGCGCGACATGCGTATTCCACCGCAGATCGACACCCCTGC 294  
 Qy 118 uLeuGluLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAs 138  
 Db 293 AGCAGCCCGATCAGCGCGCGCGCTCGATCGAGCCATAGCTGTGCTGTCGCGCGCGGAA 234  
 Qy 138 n 138  
 Db 233 T 233  
 RESULT 16  
 US-09-298-718-8/c  
 ; Sequence 8, Application US/09298718  
 ; Patent No. 6124113  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hohmann, Hans-Peter  
 ; APPLICANT: Pasamontes, Luis  
 ; APPLICANT: Tessier, Michel  
 ; APPLICANT: van Loon, Adolphus  
 ; TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION  
 ; NUMBER OF SEQUENCES: 47  
 ; CORRESPONDENCE ADDRESSES:  
 ; ADDRESSEE: Hoffmann-La Roche Inc.  
 ; STREET: 340 Kingsland Street  
 ; CITY: Nutley  
 ; STATE: NJ  
 ; COUNTRY: USA  
 ; ZIP: 07110  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; SOFTWARE: Patent in Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/298,718  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/660,645  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Pokras, Bruce A.  
 ; REGISTRATION NUMBER: 32,748  
 ; REFERENCE/DOCKET NUMBER: RAN 6002/170  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (201) 235-5801  
 ; TELEFAX: (201) 235-2363  
 ; INFORMATION FOR SEQ ID NO: 8:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1149 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; US-09-298-718-8  
 Alignment Scores:  
 Pred. No.: 8.51 Length: 1149  
 Score: 79.50 Matches: 35  
 Percent Similarity: 42.15% Conservative: 16  
 Best Local Similarity: 28.93% Mismatches: 50  
 Query Match: 11.23% Indels: 20  
 DB: 3 Gaps: 4  
 US-10-087-573-2 (1-141) x US-09-298-718-8 (1-1149)  
 Qy 37 ArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArg----- 51  
 Db 590 CGCGATCCTCGATCGAGATCGCGGTGGAGCTGAAGCGCAGCAGATAGATGAGCGGTAC 531  
 Qy 52 ---ValLeuProGlyThrGlyAlaSerAlaIleAlaIleAlaThrValThrProLysGlyAla 70

Db 530 CGTTCATCTCGGAACCGTTCGCTCATGATCATCGGCGCTCGACGCCATGGGGGGG 471  
 Qy 71 SerMetLysLeu-LysProProArg-----ProGlnSerThrLysSerProG1 86  
 Db 470 TCGGTCGTGATCTCGACGCCACCGAATTTCTGGAACCCACCGTCAGGTGCGGGGTCTCG 411  
 Qy 86 uLeuArgGluLeuSerArgLysLleArgGluMetAsnLysThrIleSerGlnGluSerAl 106  
 Db 410 A---CGCACACCGCGGTTCGATCAGCAGGAGCCTCGATCCGCGAGCGCTCCGTCAGC 354  
 Qy 106 aArgValAsnHisArg-----LeuProGluGlyHisProLe 118  
 Db 353 GTCCGCGCGGTATGTCACGCGCGACATGCGTATTCCACCGCAGATCGACACCCCTGC 294  
 Qy 118 uLeuGluLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAs 138  
 Db 293 AGCAGCCCGATCAGCGCGCGCGCTCGATCGAGCCATAGCTGTGCTGAGCGCGCGGAA 234  
 Qy 138 n 138  
 Db 233 T 233  
 RESULT 17  
 US-09-546-969-8/c  
 ; Sequence 8, Application US/09546969  
 ; Patent No. 6207409  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hohmann, Hans-Peter  
 ; APPLICANT: Pasamontes, Luis  
 ; APPLICANT: Tessier, Michel  
 ; APPLICANT: van Loon, Adolphus  
 ; TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION  
 ; NUMBER OF SEQUENCES: 47  
 ; CORRESPONDENCE ADDRESSES:  
 ; ADDRESSEE: Hoffmann-La Roche Inc.  
 ; STREET: 340 Kingsland Street  
 ; CITY: Nutley  
 ; STATE: NJ  
 ; COUNTRY: USA  
 ; ZIP: 07110  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; SOFTWARE: Patent in Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/546,969  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/660,645  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Pokras, Bruce A.  
 ; REGISTRATION NUMBER: 32,748  
 ; REFERENCE/DOCKET NUMBER: RAN 6002/170  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (201) 235-5801  
 ; TELEFAX: (201) 235-2363  
 ; INFORMATION FOR SEQ ID NO: 8:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1149 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; US-09-546-969-8  
 Alignment Scores:  
 Pred. No.: 8.51 Length: 1149  
 Score: 79.50 Matches: 35  
 Percent Similarity: 42.15% Conservative: 16



Db 7095 CCGTCCATCTCCGAAACGGTCCGCTCCATGATCATCGGCGCTCGAGCCATCGGGGGCG 7154  
 Qy 71 SerMetLysLeu-LysProProArg-----ProGlnSerThrLysSerProG1 86  
 Db 7155 TCGGTCTGATCTCGAGCCACCAATTTCTGGAACCCACGAGTCAAGTCCGGGGTCTCG 7214  
 Qy 86 uLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerA1 106  
 Db 7215 A---CGGCACCAACGGCGCTCATCAGCAGCAGCCCTCGATCCCGAGCGCTCCGTCAGC 7271  
 Qy 106 aArgValAsnHisArg-----LeuProGluGlyHisProLe 118  
 Db 7272 GTCCGGCCGGTATCGTCCAGCGTCCGAGATCGGTATTCACCCGACAGATCGACACCCCTGC 7331  
 Qy 118 uLeuGluLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAs 138  
 Db 7332 AGCAGCCGATCATCAGCGCCCGCTCGATCGAGCCATAGCTGTCTGTCAGCGCGCGAA 7391  
 Qy 138 n 138  
 Db 7392 T 7392

RESULT 20  
 US-09-452-239-13/c  
 ; Sequence 13, Application US/09452239  
 ; Patent No. 6465229  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rafalski, Antoni J.  
 ; APPLICANT: Gahoon, Rebecca E.  
 ; APPLICANT: Fader, Gary M.  
 ; TITLE OF INVENTION: Plant Caffeoyl-CoA O-Methyltransferase  
 ; FILE REFERENCE: BB1284 US NA  
 ; CURRENT APPLICATION NUMBER: US/09/452,239  
 ; EARLIER FILING DATE: 1999-12-01  
 ; EARLIER FILING DATE: 1998-December-02  
 ; NUMBER OF SEQ ID NOS: 50  
 ; SOFTWARE: Microsoft Office 97  
 ; SEQ ID NO 13  
 ; TYPE: DNA  
 ; ORGANISM: Oryza sativa  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (483)  
 US-09-452-239-13

Alignment Scores:  
 Pred. No.: 9.01 Length: 997  
 Score: 78.50 Matches: 43  
 Percent Similarity: 40.74% Conservative: 23  
 Best Local Similarity: 26.54% Mismatches: 61  
 Query Match: 11.09% Indels: 35  
 DB: 4 Gaps: 7

US-10-087-573-2 (1-141) x US-09-452-239-13 (1-997)  
 Qy 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgPro----- 16  
 Db 569 TCCGCTCCACGAACCGAAGTCGAACGCGCGCTCGGCGCCCGCGCTCTCTCGGG 510  
 Qy 17 -----ThrPheGlyGluThrPheAspValMetArgGluAlaLeuLeuArgVal 32  
 Db 509 AGCAGCTCGTCCAGCTTCTCCAGCCCTTCCAGCCCTTGGGAAGTCCACCTTGTGCGCCACC 450  
 Qy 33 -LysSerSerGluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgVa 52  
 Db 449 CCGGCTTTCTCCAGAACCGGCC---GCCGATCTCGTAGCTCTCCCTGTCCGGGTCTGATC 393  
 Qy 52 lLeuProGlyThrGlyAlaSerAlaIleAlaAlaThrValThrProLysGlyAlaSerMe 72  
 Db 392 GCCACCACTTCCGCTCTCCCGCA---GCGCAGCGCGCTCGCCAGCAGCAGTAGCGG 336

Qy 72 tLysLeuLysProPro-----ArgProGlnSerThrLysSerPro-G 86  
 Db 335 GTGAGACACCCACCTTCGATTGCTCTTCGCTCGGCCATCTTCAGCAGCATCCCCAGC 276  
 Qy 86 lLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerA 106  
 Db 275 AGCTGCGCTCATCCGCGCAGCAGTCAATGAACCCCACTGGTGTCTTCGCTGATGAGG 216  
 Qy 106 lArgValAsnHis-----ArgLeuProGluGlyHisProLeuLeuGluLysArgAlaG 124  
 Db 215 CGCAGATCGCGATGACCTCCGCTCCCGTCCGCGAGCAGC----- 178  
 Qy 124 lutyPheArgHisLeuArgSerLeuLysSerGlnGlyVal-----AsnArgLeuL 141  
 Db 177 -----CGTCTGTGCCAGGACATACCTTGTACAGGCGCTCGCTCTTGAGCAGCGTCTTG 126  
 Qy 141 le 141  
 Db 125 TT 124

RESULT 21  
 US-09-103-840A-2  
 ; Sequence 2, Application US/09103840A  
 ; Patent No. 6294328  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FLEISCHMAN, Robert D.  
 ; APPLICANT: WHITE, Owen R.  
 ; APPLICANT: FRASER, Claire M.  
 ; APPLICANT: VENTER, John C.  
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
 ; FILE REFERENCE: 24366-20007.00  
 ; CURRENT APPLICATION NUMBER: US/09/103,840A  
 ; CURRENT FILING DATE: 1998-06-24  
 ; NUMBER OF SEQ ID NOS: 2  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 2  
 ; LENGTH: 4403765  
 ; TYPE: DNA  
 ; ORGANISM: Mycobacterium tuberculosis  
 ; FEATURE:  
 ; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
 ; OTHER INFORMATION: represent a, t, c or g  
 US-09-103-840A-2

Alignment Scores:  
 Pred. No.: 7.7e+05 Length: 4403765  
 Score: 78.50 Matches: 36  
 Percent Similarity: 40.87% Conservative: 11  
 Best Local Similarity: 31.30% Mismatches: 43  
 Query Match: 11.09% Indels: 25  
 DB: 3 Gaps: 6

US-10-087-573-2 (1-141) x US-09-103-840A-2 (1-4403765)  
 Qy 41 LeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla 60  
 Db 3542568 ATCCGCGCGCTCGAAGTTCGAGTA-----CGGTAATCCCGGGTTTGTGCGAGTCGCC 3542621  
 Qy 61 lLeAlaAlaThrValThrProLysGlyAla-----SerMet 72  
 Db 3542622 GTGCGCGATGATGTCAGGTAGATCGCGGGAAGTGGAAATCGCGAGCTCAAGAAAGCTTC 3542681  
 Qy 73 LysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArg 92  
 Db 3542682 CACCTTCGCCCA-----ACCGTAGNAACCATCCGCGCAGCCAGCAGGACGTTCCGCT 3542735  
 Qy 93 LysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArg----- 107  
 Db 3542736 GACCCCGCTCCAGCAGCGCGCTATGAACGCGCGCAGCGGCGGAGCGGTGTGATGCC 3542795  
 Qy 108 -----ValAsnHisArgLeuProGluGlyHisProLeuLeu-----Glu 120

Db 3542796 CGGCCGCTCAGCGTCGTAGCCACCCCTCTCCCGAGCGGAGAACCACTCTCTGCGCTGAT 3542855  
Qy 121 LysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGln 135  
Db 3542856 GAGCGGCTCGCCGGTGGTGCAT---CGCGCGCTCGCTAGCGT 3542897  
RESULT 22  
US-09-103-840A-1  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R. R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37Rv  
US-09-103-840A-1  
Alignment Scores:  
Pred. No.: 7.71e+05 Length: 4411529  
Score: 78.50 Matches: 36  
Percent Similarity: 40.87% Conservative: 11  
Best Local Similarity: 31.30% Mismatches: 43  
Query Match: 11.09% Indels: 25  
DB: 3 Gaps: 6  
US-10-087-573-2 (1-141) x US-09-103-840A-1 (1-4411529)  
Qy 41 LeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla 60  
Db 3544986 ATCCGGCGCTCGAAGTTCAGTA-----CGGTAATCCGGGGTTTGTCCGAGTCGCC 3545039  
Qy 61 IleAlaAlaThrValThrProLysGlyAla-----SerMet 72  
Db 3545040 GTGGCCGATGTAGTCTCCAGGTAGATGCGGGGAAGTGAATCGCGAGCTCAAGAAAGCTTC 3545099  
Qy 73 LysLeuLysProProArgProGlnSerThrLysSerProLuleuArgGluLeuSerArg 92  
Db 3545100 CACCTTCGCCCA-----ACCGTAGGAACCATCCGGCCAGCCAGGAGCAACGTTCCGCT 3545153  
Qy 93 LysIleArgGluMetAnLysThrIleSerGlnGluSerAlaArg----- 107  
Db 3545154 GACCCCGTCCCGAGCGCGCGCTATGAACGCGCGAGCGCGGAACGTTGGTGGTGTATGCC 3545213  
Qy 108 -----ValaenHisArgLeuProGluGlyHisProLeuLeu-----Glu 120  
Db 3545214 CGGCCGCTCAGCGTCGTAGCCACCCCTCTCCCGAGCGGAGAACCACTCTGTGCGCTGAT 3545273  
Qy 121 LysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGln 135  
Db 3545274 GAGCGGCTCGCCGGTGGTGCAT---CGCGCGCTCGCTAGCGT 3545315  
RESULT 23  
US-09-107-532A-1348  
; Sequence 1348, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: LYNN A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310

```

; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KIRBY EADES GALE BAKER
; STREET: Box 3432, Station D
; CITY: Ottawa
; STATE: Ontario
; COUNTRY: Canada
; ZIP: K1M 1H8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/743,357
; FILING DATE: 21-AUG-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA90/00062
; FILING DATE: 23-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gale, Edwin J.
; REGISTRATION NUMBER: 28,584
; REFERENCE/DOCKET NUMBER: 30924-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 237-6900
; TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3856 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: HXB2
; US-07-743-357-20

Alignment Scores:
Pred. No.: 79.1 Length: 3856
Score: 37 Matches: 37
Percent Similarity: 40.15% Conservative: 16
Best Local Similarity: 28.03% Mismatches: 44
Query Match: 10.95% Indels: 35
DB: 2 Gaps: 6

US-10-087-573-2 (1-141) x US-07-743-357-20 (1-3856)

QY 22 PheAspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeu 41
DB 3049 TTGACTGTTTTCAGACTCTGCTATAAGA----- 3078
QY 42 ArgAlaLeuAlaGly-----MetCysGlyHisArgVal 52
DB 3079 AAGGCCTTATTAGACACATAGTACCCCTAGGTGTGATATCAACGAGACATAACAAG 3138
QY 53 LeuProGlyThrGlyAlaSerAlaIleAlaAlaThrValThrProLysGlyAlaSerMet 72
DB 3139 GTAGGATCTCTACAATACTTGGCACTAGCAGCATTAATAACACCAAAA----- 3186
QY 73 LysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArg 92
DB 3187 AGATAAAGCCACTTGTCTAGTGTAGAAACTGACAGAGATAGA----- 3234
QY 93 LysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeu 112
DB 3235 -----TGGAAACAAGCCCAAGAGACAGAGGGCCACAGAGGAGCCACACAATG 3282
QY 113 ProGluGlyHis--ProLeuLeuGluLys--ArgAlaGluTyPheArgHisLeuArgS 131

```

```

DB 3283 ---AATGACACTAGAGCTTTTAGAGAGCTTAAAGAACTGTTAGACATTTTCCTA 3339
QY 131 erLeuLysSerGlnGlyValAsnArgLeuIle 141
DB 3340 GGATTTGGCTCCATGGCTTAGGCGCAACATATC 3371

RESULT 25
US-07-743-357-21
; Sequence 21, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
; APPLICANT: Kang, Yong C.
; TITLE OF INVENTION: Polypeptide having immunological
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KIRBY EADES GALE BAKER
; STREET: Box 3432, Station D
; CITY: Ottawa
; STATE: Ontario
; COUNTRY: Canada
; ZIP: K1M 1H8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/743,357
; FILING DATE: 21-AUG-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA90/00062
; FILING DATE: 23-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gale, Edwin J.
; REGISTRATION NUMBER: 28,584
; REFERENCE/DOCKET NUMBER: 30924-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 237-6900
; TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4113 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: HXB2
; US-07-743-357-21

Alignment Scores:
Pred. No.: 86.7 Length: 4113
Score: 77.50 Matches: 37
Percent Similarity: 40.15% Conservative: 16
Best Local Similarity: 28.03% Mismatches: 44
Query Match: 10.95% Indels: 35
DB: 2 Gaps: 6

US-10-087-573-2 (1-141) x US-07-743-357-21 (1-4113)

QY 22 PheAspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeu 41
DB 3306 TTGACTGTTTTCAGACTCTGCTATAAGA----- 3335
QY 42 ArgAlaLeuAlaGly-----MetCysGlyHisArgVal 52
DB 3336 AAGGCCTTATTAGACACATAGTACCCCTAGGTGTGATATCAACGAGACATAACAAG 3395

```





```

; LOCATION: 113..1648
; OTHER INFORMATION: /product= "gag"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1408..4452
; OTHER INFORMATION: /product= "pol"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 4367..4975
; OTHER INFORMATION: /product= "sor"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 5560..8148
; OTHER INFORMATION: /product= "env"
US-08-463-210-4

Alignment Scores:
Pred. No.: 260 Length: 8933
Score: 77.50 Matches: 37
Percent Similarity: 40.15% Conservative: 16
Best Local Similarity: 28.03% Mismatches: 44
Query Match: 10.95% Indels: 35
DB: 3 Gaps: 6

US-10-087-573-2 (1-141) x US-08-463-210-4 (1-8933)
Qy 22 PheAspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeu 41
Db 4733 TTTCAGTCTTTTCAGACTCTCTATAGA-----MetCysGlyHisArgVal 52
Qy 42 ArgAlaLeuAlaGly-----MetCysGlyHisArgVal 52
Db 4763 AAGGCGCTTATTAGGACACATAGTTAGCCCTAGTGTGAATATCAAGCAGGACATAACAAG 4822
Qy 53 LeuProGlyThrGlyAlaSerAlaIleAlaAlaThrValThrProLysGlyAlaSerMet 72
Db 4823 GTAGGATCTCTACATACATTGTCAGTCTAGCAGCATTATTAACACCAAAA-----4870
Qy 73 LysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArg 92
Db 4871 AAGATAAAGCCACCTTTGCTAGTGTAGTGTAGAACTGACAGAGGATAGA-----4918
Qy 93 LysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeu 112
Db 4919 -----TGGNACAGCCCCAGAGACCAAGGCCACAGAGGCCACACACATG 4966
Qy 113 ProGluGlyHis--ProLeuLeuGluLys---ArgAlaGluTyrPheArgHisLeuArg 131
Db 4967 ---AATGGACACATAGAGCTTTTAGAGGAGCTTTAGAGATGAAGCTGTTAGACATTTCCTA 5023
Qy 131 erLeuLysSerGlnGlyValAsnArgLeu 141
Db 5024 GGAATTGGCTCCATGGCTTAGGGCAACATATC 5055

```

```

RESULT 28
US-09-620-958A-3
; Sequence 3, Application US/09620958A
; Patent No. 6294338
; GENERAL INFORMATION:
; APPLICANT: Nunomura, Kiyotada
; TITLE OF INVENTION: POLYNUCLEOTIDE AMPLIFICATION METHOD
; FILE REFERENCE: GP104-02.JT
; CURRENT APPLICATION NUMBER: US/09/620,958A
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 8933
; TYPE: RNA
; ORGANISM: Human Immunodeficiency Virus
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)...(8933)

```

```

; OTHER INFORMATION: Sequence of transcripts produced from the BH10
; Patent No. 6294338
; OTHER INFORMATION: plasmid.
US-09-620-958A-3

Alignment Scores:
Pred. No.: 260 Length: 8933
Score: 77.50 Matches: 37
Percent Similarity: 40.15% Conservative: 16
Best Local Similarity: 28.03% Mismatches: 44
Query Match: 10.95% Indels: 35
DB: 3 Gaps: 6

US-10-087-573-2 (1-141) x US-09-620-958A-3 (1-8933)
Qy 22 PheAspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeu 41
Db 4733 UUUGACUUUUUUCAGACUCUCUAUAGA-----MetCysGlyHisArgVal 52
Qy 42 ArgAlaLeuAlaGly-----MetCysGlyHisArgVal 52
Db 4763 AAGCCUUUUUAGGACACACAUUAGCCUUGGUGUGUAUAUACAAGCAGGACACAAAG 4822
Qy 53 LeuProGlyThrGlyAlaSerAlaIleAlaAlaThrValThrProLysGlyAlaSerMet 72
Db 4823 GUAGGAUCUCUACAUAUACUUGGCACUAGUAGCCUUGGUGUGUAUAUACAAGCAGGACACAAAG 4870
Qy 73 LysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArg 92
Db 4871 AAGAUAAAGCCACCUUUGCUAGUGUUAUACGAACUCAGAGGGAUAGA-----4918
Qy 93 LysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeu 112
Db 4919 -----UGGAACAGCCCCAGAGACCAAGGCCACAGAGGCCACACAAUG 4966
Qy 113 ProGluGlyHis--ProLeuLeuGluLys---ArgAlaGluTyrPheArgHisLeuArg 131
Db 4967 ---AUGGACACUAGAGCUUUUAGAGGCUUAAGAAGCUGUUAAGCAUUUCCUA 5023
Qy 131 erLeuLysSerGlnGlyValAsnArgLeu 141
Db 5024 GGAUUUGCCUCCAUUGGCUUAGGCAACAUAUC 5055

RESULT 29
US-09-620-958A-4
; Sequence 4, Application US/09620958A
; Patent No. 6294338
; GENERAL INFORMATION:
; APPLICANT: Nunomura, Kiyotada
; TITLE OF INVENTION: POLYNUCLEOTIDE AMPLIFICATION METHOD
; FILE REFERENCE: GP104-02.JT
; CURRENT APPLICATION NUMBER: US/09/620,958A
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 8933
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the IAC-Asrc pseudo target
; Patent No. 6294338
; NAME/KEY: mutation
; LOCATION: (4135)...(4155)
; OTHER INFORMATION: Mutated positions: 4135, 4140-1, 4145, 4150,
; OTHER INFORMATION: 4152-3, 4155
US-09-620-958A-4

Alignment Scores:
Pred. No.: 260 Length: 8933
Score: 77.50 Matches: 37
Percent Similarity: 40.15% Conservative: 16
Best Local Similarity: 28.03% Mismatches: 44

```

```

Query Match: 10.95% 35
DB: Gaps: 6

US-10-087-573-2 (1-141) x US-09-620-958A-4 (1-8933)

QY 22 PheAspValMetArgGluAlaLeuLeuLeuArgValLysSerSerGluArgLeuAlaMetLeu 41
DB 4733 UUGACUGUUUUUCACACUCUGCAUAAGA-----4762
QY 42 ArgAlaLeuAlaGly-----MetCysGlyHisArgVal 52
DB 4763 AAGCCCUUAUAGGACACAUAGUAGCCUAGGUGUAUAUCAACAGCAGCAUAACAAG 4822
QY 53 LeuProGlyThrGlyAlaSerAlaIleAlaAlaThrValThrProLysGlyAlaSerMet 72
DB 4823 GUAGGAUCUACAUAUCUUGGCACUAGCAGCAUAUAUAACACCAAAA-----4870
QY 73 LysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArg 92
DB 4871 AAGAUAAGCCACCUUUGCUAGUGUUGACAAACUCAGACAGAGGAUAGA-----4918
QY 93 LysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeu 112
DB 4919 -----UGGAACAGCCCCACAGACACCAAGGCCACAGAGGGGCCACACAUG 4966
QY 113 ProGluGlyHis--ProLeuLeuGluLys---ArgAlaGluTyrPheArgHisLeuArgS 131
DB 4967 ---AAUGCACACUAGAGCUUUUAGAGGCUUAAGAAUGAAGCGUGUAGACAUUUUCCUA 5023
QY 131 erLeuLysSerGlnGlyValAsnArgLeuIle 141
DB 5024 GGAUUUGGCUCCAUUGCUUAGGGCAACAUAUC 5055

RESULT 30
US-09-620-958A-9
; Sequence 9, Application US/09620958A
; Patent No. 6294338
; GENERAL INFORMATION:
; APPLICANT: Nunomura, Kiyotada
; TITLE OF INVENTION: POLYNUCLEOTIDE AMPLIFICATION METHOD
; FILE REFERENCE: GP104-02.UT
; CURRENT APPLICATION NUMBER: US/09/620,958A
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 8933
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the IAC-Bscr pseudo target
; Patent No. 6294338
; NAME/KEY: mutation
; LOCATION: (4140)...(4159)
; OTHER INFORMATION: Mutated positions: 4140-42, 4145-47, 4152,
; OTHER INFORMATION: 4156-57, 4159
US-09-620-958A-9

Alignment Scores:
Pred. No.: 260 Length: 8933
Score: 77.50 Matches: 37
Percent Similarity: 40.15% Conservative: 16
Best Local Similarity: 28.03% Mismatch: 44
Query Match: 10.95% Indels: 35
DB: Gaps: 6

US-10-087-573-2 (1-141) x US-09-620-958A-9 (1-8933)

QY 22 PheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeu 41
DB 4733 UUGACUGUUUUUCACACUCUGCAUAAGA-----4762
QY 42 ArgAlaLeuAlaGly-----MetCysGlyHisArgVal 52

```

Db 5560 -----TGGAAACAGCCCAAGAGACCAAGGGCCACAGAGGAGGACCAACAATG 5607  
QY 113 ProGluGlyHis--ProLeuLeuGluLys--ArgAlaGluTyrPheArgHisLeuArgS 131  
Db 5608 ----AATGGACACTAGAGCTTTTAGAGAGCTTAAGAACTGTTAGACATTTTCCTA 5664  
QY 131 erLeuLysSerGlnGlyValAsnArgLeu141  
Db 5665 GGATTGGCTCCATGCTTAGGGCAACATATC 5696

## RESULT 32

US-09-441-340-2  
; Sequence 2, Application US/09441340  
; Patent No. 6448476  
; GENERAL INFORMATION:  
; APPLICANT: Barry, Gerard F.  
; TITLE OF INVENTION: Phosphonate Metabolizing Plants  
; FILE REFERENCE: 38-21(15303)  
; CURRENT APPLICATION NUMBER: US/09/441,340  
; EARLIER FILING DATE: 1999-11-16  
; EARLIER FILING DATE: 1998-11-17  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 11672  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
US-09-441-340-2

## Alignment Scores:

Pred. No.: 379 Length: 11672  
Score: 77.50 Matches: 35  
Percent Similarity: 39.72% Conservative: 21  
Best Local Similarity: 24.82% Mismatches: 55  
Query Match: 10.95% Indels: 30  
DB: 4 Gaps: 4

US-10-087-573-2 (1-141) x US-09-441-340-2 (1-11672)

QY 7 ThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg 26  
Db 8536 ACCACTCGCGGGCCAGCCAGTTCGCAACCGCGAGAGATTCGGGAATATTATCAGG 8595  
QY 27 -----GluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArgAla 43  
Db 8596 GCAAACTACTCCTCACTGATCGGCAGATGCAGCAGTACGAAGAAGAGCAACTGGCGCTCG 8655  
QY 44 LeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAlaIleAlaAla 63  
Db 8656 -----CCGACCGCTGTCGACCGCGAATCGCGAAT 8685  
QY 64 ThrValThrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSerThrLys 83  
Db 8686 CCATCGCGCCCTGTGCGCGCGGAAAAATTTGGCTTGCAGCCAGATGACGCCACCC 8745  
QY 84 SerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThr-IleSerGln 103  
Db 8746 ACGCCCACTTG-----CTGAATCTCACCAGCTTGCA 8778  
QY 103 nGluSerAlaArgValAsnHisArgLeuProGluGly-----Hi 116  
Db 8779 GCGTGATCGCGAATTTCCACCACTGTCGAAGCGCGGAAGCTCGCGCAAGCATGGCA 8838  
QY 116 sProLeuLeuLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGln 136  
Db 8839 TGAACGTGCTGATGGCGCGCGAATATTGTGCGCGCGGCTCGCACTCCGCAACGTGG 8898  
QY 136 Y 136  
Db 8899 C 8899

## RESULT 33

US-09-441-340-1  
; Sequence 1, Application US/09441340  
; Patent No. 6448476  
; GENERAL INFORMATION:  
; APPLICANT: Barry, Gerard F.  
; TITLE OF INVENTION: Phosphonate Metabolizing Plants  
; FILE REFERENCE: 38-21(15303)  
; CURRENT APPLICATION NUMBER: US/09/441,340  
; EARLIER FILING DATE: 1999-11-16  
; EARLIER FILING DATE: 1998-11-17  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 15611  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
US-09-441-340-1

## Alignment Scores:

Pred. No.: 572 Length: 15611  
Score: 77.50 Matches: 35  
Percent Similarity: 39.72% Conservative: 21  
Best Local Similarity: 24.82% Mismatches: 55  
Query Match: 10.95% Indels: 30  
DB: 4 Gaps: 4

US-10-087-573-2 (1-141) x US-09-441-340-1 (1-15611)

QY 7 ThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg 26  
Db 12772 ACCACTCGCGGGCCAGCCAGTTCGCAACCGCGAGAGATTCGGGAATATTATCAGG 12831  
QY 27 -----GluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArgAla 43  
Db 12832 GCAAACTACTCCTCACTGATCGGCAGATGCAGCAGTACGAAGAAGAGCAACTGGCGCTCG 12891  
QY 44 LeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAlaIleAlaAla 63  
Db 12892 -----CCGACCGCTGTCGCGAATCGCGAAT 12921  
QY 64 ThrValThrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSerThrLys 83  
Db 12922 CCATCGCGCCCTGTGCGCGCGGAAAAATTTGGCTTGCAGCCAGATGACGCCACCC 12981  
QY 84 SerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThr-IleSerGln 103  
Db 12982 ACGCCCACTTG-----CTGAATCTCACCAGCTTGCA 13014  
QY 103 nGluSerAlaArgValAsnHisArgLeuProGluGly-----Hi 116  
Db 13015 GCGTGATCGCGAATTTCCACCACTGTCGAAGCGCGGAAGCTCGCGCAAGCATGGCA 13074  
QY 116 sProLeuLeuLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGln 136  
Db 13075 TGAACGTGCTGATGGCGCGCGAATATTGTGCGCGCGGCTCGCACTCCGCAACGTGG 13134  
QY 136 Y 136  
Db 13135 C 13135

## RESULT 34

US-07-977-434-7  
; Sequence 7, Application US/07977434  
; Patent No. 5466591  
; GENERAL INFORMATION:  
; APPLICANT: Gelfand, David H.  
; APPLICANT: Abramson, Richard D.  
; TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF  
; THERMOSTABLE DNA POLYMERASES  
; NUMBER OF SEQUENCES: 38  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann-La Roche Inc.  
STREET: 340 Kingsland Street  
CITY: Nutley  
STATE: New Jersey  
ZIP: 07110-1199  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: 7  
SOFTWARE: Wordperfect 2.1  
CURRENT APPLICATION DATA:  
APPLYING DATE: US/07/977,434  
FILING DATE: US/07/977,434  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 590,490  
FILING DATE: 28-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 590,466  
FILING DATE: 28-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 590,213  
FILING DATE: 28-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 523,394  
FILING DATE: 15-MAY-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 143,441  
FILING DATE: 12-JAN-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 063,509  
FILING DATE: 17-JUN-1987  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 899,241  
FILING DATE: 22-AUG-1986  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 746,121  
FILING DATE: 15-AUG-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US90/07641  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 585,471  
FILING DATE: 20-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 455,611  
FILING DATE: 22-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 609,157  
FILING DATE: 02-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 557,517  
FILING DATE: 24-JUL-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Luann Cserr  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: Case No. 5466591 8753  
TELEPHONE: (510) 814-2972  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2505 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Thermus species 205  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..2502

US-07-977-434-7

Alignment Scores:

Pred. NO.: 48.9 Length: 2505  
Score: 77.00 Matches: 53  
Percent Similarity: 41.48% Conservative: 20  
Best Local Similarity: 30.11% Mismatches: 62  
Query Match: 10.88% Indels: 42  
DB: 1 Gaps: 9

US-10-087-573-2 (1-141) x US-07-977-434-7 (1-2505)

QY 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20  
DB 1144 CTGGACCCCTCCACACACCCCGAGGGGTGGCC-----CGGCGCTACGGGGGGGAG 1197  
QY 21 ---ThrPheAspValMetArgGluAlaLeuLeu----- 30  
DB 1198 TGGACGGAGGACCGCCCGCCCGGGGCTCTCTCGCCGAGCGGCTCCAGCAAAACCTTTG 1257  
QY 31 ---ArgValLysSerSerGluArgLeuAlaMetLeu----- 41  
DB 1258 GAACGCTCAGGGAGAGGAAAGCTCTTTGGCTCTACCAAGAGGTGGAAGCCCTC 1317  
QY 42 ---ArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSer--- 59  
DB 1318 TCCCGGGTCTGGCCACATGGA-GGCCACCGGGTAAGCTGACGCTGCTATCTAAA 1376  
QY 60 -----AlaIleAlaAlaThrValThrProLysGlyAlaSerMet----- 72  
DB 1377 GCGCCTTTCCCTGGAGCTTGGCGAGGAGATTCGCGGCTCGAGGAGGAGGTCTTCGCGCT 1436  
QY 73 -----LysLeuLysProProArgProGlnSerThrLysSerProGluLeu-ArgGl 89  
DB 1437 GGCGGGCCACCCCTTCAACCTGAACTCCCGTGACAGCTAGAGCGGTGCTCTTTGACGA 1496  
QY 89 uLeuSerArgLysIleArgGluMetAsnLysThrIleSer-----GlnGluSe 105  
DB 1497 GCTT-----AGGCTTCCCGCCCTGGGCAAGACGCAAAACGCGGGAGCGCTCCACAG 1550  
QY 105 rAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGluTy 125  
DB 1551 CGCCCGGTGCTGGAGGCCCTCAGGGAGGCCACCCCATCGTGAGAGAAGATCCTCCAGCA 1610  
QY 125 rPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140  
DB 1611 C---CGGAGCTCACCAGCTCAAGAACACACCTACGTGGACCCCTC 1653

RESULT 35

US-08-458-819-7  
; Sequence 7, Application US/08458819  
; Patent No. 5795762  
; GENERAL INFORMATION:  
; APPLICANT: Gelfand, David H.  
; APPLICANT: Abramson, Richard D.  
; TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF  
; TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES  
; NUMBER OF SEQUENCES: 38  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hoffmann-La Roche Inc.  
; STREET: 340 Kingsland Street  
; CITY: Nutley  
; STATE: New Jersey  
; ZIP: 07110-1199  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: 7  
; SOFTWARE: Wordperfect 2.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/458,819  
; FILING DATE: 02-JUN-1995  
; CLASSIFICATION: 435

QY	1	MetGluSerThrSerThrThrThrAenPheValAlaGluAAsnArgProThrPheGlyGlu	20
Db	1144	CTGACCCCTTCAACACACCCCGAGGGGTGGCC-----CGGCGCTACGGGGGGAG	1197
QY	21	---ThrPheAspValMetArgGluAlaLeuLeu-----	30
Db	1198	TGGACGGAGGACCGCCGCCACCGGGCCCTCTCGCCGAGCGGTCCAGCAAAACCTCTTG	1257
QY	31	---ArgVallySerSerGluArgLeuAlaMetLeu-----	41
Db	1358	GAACGCTCAAGGAGGAGGAAAGCTCTTTGGCTTACCAAGAGGTGGAAAGCCCTC	1317
QY	42	---ArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSer---	59
Db	1318	TCCCGGGTCTCGCCACATGGA-GGCCACCGGGTAAGCTGCGCTATCTATAA	1376
QY	60	-----AlaIleAlaAlaThrValThrProLysGlyAlaSerMet-----	72
Db	1377	GGCCCTTCCCTGGAGCTTCGGAGAGAGATTCCCGCCCTCGAGAGAGAGGTCTTCGCGCT	1436
QY	73	-----LysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGlu	89
Db	1437	GGCGGGCCACCCCTTCAACCTGAACCTCCCTGACAGCTAGAGCGGTGCTCTTTGACGA	1496
QY	89	uLeuSerArgLysIleArgGluMetAsnLysThrIleSer-----GlnGlu	105
Db	1497	GCCTT-----AGGCTTCCCGCTGGGCAAGAGCGAAAGACGGGAGCGCTCCACCG	1550
QY	105	rAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGluTy	125
Db	1551	CGCGCGGTGCTGGAGGCCCTCTCAGGGAGGCCACCCCATCGTGAGAAGATCCTCCAGCA	1610
QY	125	rPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu	140
Db	1611	C---CGGAGCTCACCAGCTCAAGACACTAGTGACCCCTC	1653
RESULT 36			
PCT-US91-07035-7			
Sequence 7, Application PC/TUS9107035			
GENERAL INFORMATION:			
APPLICANT: Gelfand, David H.			
APPLICANT: Abramson, Richard D.			
TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF			
TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES			
NUMBER OF SEQUENCES: 38			
CORRESPONDENCE ADDRESS:			
ADDRESS: Cetus Corporation			
STREET: 1400 Fifty-third Street			
CITY: Emeryville			
STATE: California			
ZIP: 94608			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: Wordperfect 5.0			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: PCT/US91/07035			
FILING DATE: 19910930			
CLASSIFICATION: 435			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 590,490			
FILING DATE: 28-SEP-1990			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 590,466			
FILING DATE: 28-SEP-1990			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 590,213			
FILING DATE: 28-SEP-1990			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 523,394			
FILING DATE: 15-MAY-1990			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 523,394			
FILING DATE: 15-MAY-1990			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 523,394			
FILING DATE: 15-MAY-1990			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 523,394			
FILING DATE: 15-MAY-1990			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 523,394			
FILING DATE: 15-MAY-1990			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 523,394			
FILING DATE: 15-MAY-1990			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 523,394			
FILING DATE: 15-MAY-1990			
PRIOR APPLICATION DATA:			

Alignment Scores:		
Pred. No.:	48-9	Length:
Score:	77.00	Matches:
Percent Similarity:	41.48%	Conservative:
Best local Similarity:	30.11%	Mismatches:
Query Match:	10.88%	Indels:
DB:	1	Gaps:
		9
US-10-C87-573-2 (1-141)	x US-08-458-819-7 (1-250S)	

APPLICATION NUMBER: US 143,441  
 FILING DATE: 12-JAN-1988  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 063,509  
 FILING DATE: 17-JUN-1987  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 899,241  
 FILING DATE: 22-AUG-1986  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 746,121  
 FILING DATE: 15-AUG-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: WO PCT/US90/07641  
 FILING DATE: 21-DEC-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 585,471  
 FILING DATE: 20-SEP-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 455,611  
 FILING DATE: 22-DEC-1989  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 609,157  
 FILING DATE: 02-NOV-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 557,517  
 FILING DATE: 24-JUL-1990  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Sias Ph.D. Stacey R.  
 REGISTRATION NUMBER: 32,630  
 REFERENCE/DOCKET NUMBER: Case No. 2580  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-420-3300  
 INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2505 base pairs  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Thermus species Z05  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..2502  
 PCT-US91-07035-7

Alignment Scores:  
 Pred. No.: 48.9 Length: 2505  
 Score: 77.00 Matches: 53  
 Percent Similarity: 41.48% Conservative: 20  
 Best Local Similarity: 30.11% Mismatches: 62  
 Query Match: 10.88% Indels: 42  
 DB: 5 Gaps: 9

US-10-087-573-2 (1-141) x PCT-US91-07035-7 (1-2505)  
 QY 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20  
 Db 1144 CTGGACCCCTCCACACACACACCCCGAGGGGTGCCC-----CGGCGCTACGGGGGGGAG 1197  
 QY 21 ---ThrPheAspValMetArgGluAlaLeuLeu----- 30  
 Db 1198 TGGACGGAGGACGGCCCGCCCGGCGCTCTCCCGGAGGCTCCAGCAAAACCTCTTG 1257  
 QY 31 ---ArgValLysSerSerGluArgLeuAlaMetLeu----- 41  
 Db 1258 GAACGCTCAAGGGAGGAGGAAAGCTCTTTGGTCTTACCAAGAGGTGGAAAGCCCGCTC 1317  
 QY 42 ---ArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSer--- 59  
 Db 1318 TCCGGGTCTTGGCCCATGGA-GGCCACCGGGGTAGGCTGACGCTGAGCTTCTATAA 1376

QY 60 -----AlaIleAlaThrValThrProLysGlyAlaSerMet----- 72  
 Db 1377 GGCCCTTTCCCTGGAGCTTGGGAGGAGATTCCGCGCTCCAGGAGGAGTCTTCCGCT 1436  
 QY 73 -----LysLeuLysProArgProGlnSerThrLysSerProGluLeu-ArgG1 89  
 Db 1437 GCGGGCCACCCCTTCAACCTGAACCTCCGTCGACGAGTACGCGGGTCTTTTGACGA 1496  
 QY 89 uleuSerArgLysIleArgGluMetAsnLysThrIleSer-----GlnGluSe 105  
 Db 1497 GCTT-----AGGCTTCCCGCTTGGCAAGACGCAAAAGACGGGAAAGCGCTCCACCAG 1550  
 QY 105 rAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGluTyr 125  
 Db 1551 GCGCGCGGTGCTGGAGGCGCTCAGGAGGCGCCACCCCATCTGTGGAGAGATCCTCCAGCA 1610  
 QY 125 rPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140  
 Db 1611 C---CGGAGCTCACCAAGCTCAAGACACACCTACGTGACCCCTC 1653

RESULT 37  
 US-09-335-409-1/c  
 ; Sequence 1, Application US/09335409  
 ; Patent No. 6121029  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schupp, Thomas  
 ; APPLICANT: Ligon, James  
 ; APPLICANT: Molnar, Istvan  
 ; APPLICANT: Zirkle, Ross  
 ; APPLICANT: Cyr, Devon  
 ; APPLICANT: Goerlach, Joern  
 ; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
 ; FILE REFERENCE: 4-30582A  
 ; CURRENT APPLICATION NUMBER: US/09/335,409  
 ; CURRENT FILING DATE: 1999-06-17  
 ; NUMBER OF SEQ ID NOS: 30  
 ; SOFTWARE: Patent in Ver. 2.0  
 ; SEQ ID NO 1  
 ; LENGTH: 68750  
 ; TYPE: DNA  
 ; ORGANISM: Sorangium cellulosum  
 ; US-09-335-409-1

Alignment Scores:  
 Pred. No.: 5.3e+03 Length: 68750  
 Score: 77.00 Matches: 34  
 Percent Similarity: 45.19% Conservative: 13  
 Best Local Similarity: 32.69% Mismatches: 28  
 Query Match: 10.88% Indels: 29  
 DB: 3 Gaps: 5

US-10-087-573-2 (1-141) x US-09-335-409-1 (1-68750)  
 QY 8 ThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArgGlu 27  
 Db 54914 ACCAAGATCTGCGGAGCTGCGCCAGCTTTTCGGGAG-----CAACGC 54870  
 QY 28 AlaLeuLeuArgValLysSerSerGluArgLeu----- 38  
 Db 54869 GTCTTCTCTGCTCGCTCATGCCCCGCGAGAGCTTCGAGATCGCGGATCGTTCTCGAA 54810  
 QY 39 AlaMetLeu-ArgAlaLeuAlaGly-----MetCysGlyHisArgVal-- 52  
 Db 54809 GCTTCTTCTCCGCTCGGTGCGCGGAGCGCTGGGTGCGCGCGGATTCGGAACAGATGTC 54750  
 QY 53 -----LeuProGlyThrGlyAla-----SerAlaIleAlaAlaTh 64  
 Db 54749 TAGCAAGCTCCCGCTCAGAGCTGCTACGTAGGTAGGTCCATAGCAGGCTCGCGGCAC 54690  
 QY 64 rValThrProLysGlyAlaSerMetLysLeuLysProArgProGlnSerThrLysSe 84  
 Db 54689 GGTATGATCCGAGCGCGGCTTCATGATCGGTTCGGAGCTTCAGGCTATCAGGAGTCCAT 54630



## Alignment Scores:

Pred. No.:	5.3e+03	Length:	68750
Score:	77.00	Matches:	34
Percent Similarity:	45.19%	Conservative:	13
Best Local Similarity:	32.69%	Mismatches:	28
Query Match:	10.88%	Indels:	29
DB:	4	Gaps:	5

US-10-087-573-2 (1-141) x US-09-568-480-1 (1-68750)

Qy	8	ThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArgGlu	27
Db	54914	ACCAACGATCTCGCGAGCTGCGCCAGCTTTTCGGCGAG-----CAACGC	54870
Qy	28	AlaLeuLeuArgValLysSerSerGluArgLeu-----	38
Db	54869	GTCCCTCTGCTCGTCCGTCATGCCCCGAGAGCCTCGAGATCTCGGGCATCGTTCTCGAA	54810
Qy	39	AlaMetLeu-ArgAlaLeuAlaGly-----MetCysGlyHisArgVal--	52
Db	54809	GCTCTCTCCCGCTCGGTGCGCGGAGCGTGGTCCGCCGGCATTTCGGAAACAGAAATGTC	54750
Qy	53	-----LeuProGlyThrGlyAla-----SerAlaIleAlaAlaTh	64
Db	54749	TAGCAAGCTCCGCTCAGAGCTGCTACGTTAGGGTAGGTGTCATAGCAGGGTCCGGCGCAC	54690
Qy	64	rValThrProLysGlyAlaSerMetLysLeuLysProArgProGlnSerThrLysSe	84
Db	54689	GGTGATGCCGAGCGCGCTCGATCGGTTSCGGAGCTCCAGGCCCTATCAGCGAGTCCAT	54630
Qy	84	rProGluLeu	87
Db	54629	GCCGAGATTG	54620

Search completed: November 17, 2003, 14:38:38

Job time : 1606 secs



OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 17, 2003, 13:45:59 ; Search time 241 Seconds  
(without alignments)  
1912.103 Million cell updates/sec

Title: US-10-087-573-2

Perfect score: 708

Sequence: 1 MESTSTTTFVAENRPTFGE.....RAEYFRHLRLSLKSGVNRLLI 141

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh

-Q=/cgn2\_1/USPTO.spool/US10087573/runat\_14112003\_103612\_20787/app\_query.fasta\_1.327

-DB=Published Applications NA -QFMT=fastap -SUFFIX=nb -MINWATCH=0.1

-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62

-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct.THR MAX=100

-THR\_MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HAPSIZ=500 -MINLEN=0

-MAXLEN=2000000000 -USER=US10087573 @CGN\_1\_1\_164 @runat\_14112003\_103612\_20787

-NCPU=6 -ICPU=3 -NO\_WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100

-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5

-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*

2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*

3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*

4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*

5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*

6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*

7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*

8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*

9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*

10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*

11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*

12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*

13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*

14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*

15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*

16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*

17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Length	ID	Description
1	708	100.0	1135	12 US-10-087-573-1

2	680	96.0	1134	12	US-10-087-573-3	Sequence 3, Appli
3	92.5	13.1	10531	14	US-10-160-758-10	Sequence 10, Appli
4	92	13.0	90	12	US-10-087-573-5	Sequence 5, Appli
5	90	12.7	1659	10	US-09-796-692-667	Sequence 667, App
6	90	12.7	1659	14	US-10-040-862-667	Sequence 667, App
7	88.5	12.5	437	11	US-09-918-995-36271	Sequence 36271, A
8	88.5	12.5	442	11	US-09-918-995-36013	Sequence 36013, A
9	85	12.0	486	11	US-09-918-995-8937	Sequence 8937, Ap
10	84	11.9	2517	11	US-09-864-636A-89	Sequence 89, Appli
11	84	11.9	2517	12	US-09-758-282-120	Sequence 120, App
12	84	11.9	2517	12	US-10-084-839-89	Sequence 89, Appli
13	82.5	11.7	505	12	US-10-029-386-5420	Sequence 5420, Ap
14	82.5	11.7	747	14	US-10-156-761-7288	Sequence 7288, Ap
15	82.5	11.7	9025608	14	US-10-156-761-1	Sequence 1, Appli
16	82	11.6	786	14	US-10-156-761-3722	Sequence 3722, Ap
17	82	11.6	9025608	14	US-10-156-761-1	Sequence 1, Appli
18	81.5	11.5	9718	8	US-08-319-974A-1	Sequence 1, Appli
19	81	11.4	1797	9	US-09-833-102-1	Sequence 1, Appli
20	81	11.4	1797	9	US-09-832-441-1	Sequence 1, Appli
21	81	11.4	1902	14	US-10-156-761-3584	Sequence 2584, Ap
22	81	11.4	2517	12	US-10-084-839-2690	Sequence 2690, Ap
23	81	11.4	3530	10	US-09-969-708-93	Sequence 93, Appli
24	81	11.4	3530	10	US-09-969-708-822	Sequence 622, App
25	81	11.4	3530	10	US-09-880-107-3771	Sequence 3771, Ap
26	80.5	11.4	993	14	US-10-156-761-1190	Sequence 1190, Ap
27	80	11.3	2517	11	US-09-864-636A-96	Sequence 96, Appli
28	80	11.3	2517	12	US-09-758-282-141	Sequence 141, App
29	80	11.3	2517	12	US-10-084-839-96	Sequence 96, Appli
30	79.5	11.2	1149	10	US-09-547-287-8	Sequence 8, Appli
31	79.5	11.2	1668	14	US-10-156-761-2702	Sequence 2702, Ap
32	79.5	11.2	8625	11	US-09-920-923-1	Sequence 1, Appli
33	79.5	11.2	11233	11	US-09-920-923-27	Sequence 27, Appli
34	79	11.2	756	14	US-10-156-761-6180	Sequence 6180, Ap
35	79	11.2	936	14	US-10-156-761-6333	Sequence 6333, Ap
36	79	11.2	1860	14	US-10-156-761-1558	Sequence 1558, Ap
37	79	11.2	2508	11	US-09-864-636A-431	Sequence 431, App
38	79	11.2	2508	12	US-10-084-839-431	Sequence 431, App
39	79	11.2	2517	11	US-09-864-636A-73	Sequence 73, Appli
40	79	11.2	2517	11	US-09-864-636A-87	Sequence 87, Appli
41	79	11.2	2517	11	US-09-864-636A-88	Sequence 88, Appli
42	79	11.2	2517	11	US-09-864-636A-91	Sequence 91, Appli
43	79	11.2	2517	11	US-09-864-636A-92	Sequence 92, Appli
44	79	11.2	2517	11	US-09-864-636A-93	Sequence 93, Appli
45	79	11.2	2517	11	US-09-864-636A-97	Sequence 97, Appli

#### ALIGNMENTS

RESULT 1

US-10-087-573-1

; Sequence 1, Application US/10087573

; Publication No. US20030165872A1

; GENERAL INFORMATION:

; APPLICANT: SCHEITERS, Theodorus PM

; APPLICANT: CARCY, Bernard PD

; APPLICANT: DRACULOVSKI, Pascal R

; APPLICANT: GORENFLOT, Andre F

; TITLE OF INVENTION: BABESIA CANIS VACINE

; FILE REFERENCE: SCHEITERS

; CURRENT APPLICATION NUMBER: US/10/087,573

; PRIOR FILING DATE: 2002-02-28

; PRIOR FILING DATE: 2001-03-06

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 1135

; TYPE: DNA

; ORGANISM: Babesia canis

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (75)..(500)

US-10-087-573-1

Alignment Scores:  
Pred. No.: 2,04e-77 Length: 1135  
Score: 708.00 Matches: 141  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 12 Gaps: 0

US-10-087-573-2 (1-141) x US-10-087-573-1 (1-1135)

Qy 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20  
Db 75 ATGGAGTCGACATCAACACCAACCACTTTGTCCGAGAACCGTCCACCTTTGGTGAG 134

Qy 21 ThrPheAspValMetArgGluAlaLeuLeuValLysSerSerGluArgLeuAlaMet 40  
Db 135 ACGTTTGATGATGAGGAGGAGCTTCTCGTGTAAAGTCTCTGAACGCTTGGCAATG 194

Qy 41 LeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla 60  
Db 195 CTCAGAGCGCTTGCAGGAATGTCGGTCCACCGCTCTTCTGCGACTGGTCTCTGCG 254

Qy 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLysProProArgProGln 80  
Db 255 ATAGCGGCAACGGTAACCCCAAGGGGCTTCGATGAAGCTTAAACCAACCGCTCCGAG 314

Qy 81 SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThr 100  
Db 315 TCAACGAAAGTCTCCGAGCTCAGGAGCTGTCACGGAAGTTCGCGAAATGAATAAGACT 374

Qy 101 IleSerGlnGluSerAlaArgValAenHisArgLeuProGluGlyHisProLeuLeuGlu 120  
Db 375 ATAAGTCAGGAATCAGCTCGGTAAACCAACCGCTTCCGGAAGCCACCTCTCTTAGAG 434

Qy 121 LysArgAlaGluTyPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140  
Db 435 AAGCGGGCAGAGATATTTTCGTACCTTAGATCTCTTAAAGCCCAAGGAGTCAATAGACTC 494

Qy 141 Ile 141  
Db 495 ATC 497

RESULT 2  
US-10-087-573-3  
; Sequence 3, Application US/10087573  
; Publication No. US20030165872A1  
; GENERAL INFORMATION:  
; APPLICANT: SCHEITERS, Theodorius PM  
; APPLICANT: CARCY, Bernard PD  
; APPLICANT: DRACULOVSKI, Pascal R  
; APPLICANT: GORENFLOT, Andre F  
; TITLE OF INVENTION: BABESIA CANIS VACINE  
; FILE REFERENCE: SCHEITERS  
; CURRENT APPLICATION NUMBER: US/10/087,573  
; CURRENT FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: EP 01200816.5  
; PRIOR FILING DATE: 2001-03-06  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 1134  
; TYPE: DNA  
; ORGANISM: Babesia canis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (75) .. (929)  
US-10-087-573-3

Alignment Scores:  
Pred. No.: 5,67e-74 Length: 1134  
Score: 680.00 Matches: 140  
Percent Similarity: 99.29% Conservative: 0

Best Local Similarity: 99.29% Mismatches: 1  
Query Match: 96.05% Indels: 1  
DB: 12 Gaps: 0

US-10-087-573-2 (1-141) x US-10-087-573-3 (1-1134)

Qy 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20  
Db 75 ATGGAGTCGACATCAACACCAACCACTTTGTCCGAGAACCGTCCACCTTTGGTGAG 134

Qy 21 ThrPheAspValMetArgGluAlaLeuLeuValLysSerSerGluArgLeuAlaMet 40  
Db 135 ACGTTTGATGATGAGGAGGAGCTTCTCGTGTAAAGTCTCTGAACGCTTGGCAATG 194

Qy 41 LeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla 60  
Db 195 CTCAGAGCGCTTGCAGGAATGTCGGTCCACCGCTCTTCTGCGACTGGTCTCTGCG 254

Qy 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLysProProArgProGln 80  
Db 255 ATAGCGGCAACGGTAACCCCAAGGGGCTTCGATGAAGCTTAAACCAACCGCTCCGAG 314

Qy 81 SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThr 100  
Db 315 TCAACGAAAGTCTCCGAGCTCAGGAGCTGTCACGGAAGTTCGCGAAATGAATAAGACT 374

Qy 101 IleSerGlnGluSerAlaArgValAenHisArgLeuProGluGlyHisProLeuLeuGlu 120  
Db 375 ATAAGTCAGGAATCAGCTCGGTAAACCAACCGCTTCCGGAAGCCACCTCTCTTAGAG 434

Qy 121 LysArgAlaGluTyPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140  
Db 435 AAGCGGGCAGAGATATTTTCGTACCTTAGATCTCTTAAAGCCCAAGGAGTCAATAGACTC 493

Qy 141 Ile 141  
Db 494 ATC 496

RESULT 3  
US-10-160-758-10/c  
; Sequence 10, Application US/10160758  
; Publication No. US20030036076A1  
; GENERAL INFORMATION:  
; APPLICANT: EXELIXIS, INC.  
; TITLE OF INVENTION: CADS AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE  
; FILE REFERENCE: EX02-089C  
; CURRENT APPLICATION NUMBER: US/10/160,758  
; CURRENT FILING DATE: 2002-06-03  
; PRIOR APPLICATION NUMBER: US 60/296,076  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/328,605  
; PRIOR FILING DATE: 2001-10-10  
; PRIOR APPLICATION NUMBER: US 60/357,253  
; PRIOR FILING DATE: 2002-02-15  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 10531  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-160-758-10

Alignment Scores:  
Pred. No.: 2.1 Length: 10531  
Score: 92.50 Matches: 38  
Percent Similarity: 46.58% Conservative: 30  
Best Local Similarity: 26.03% Mismatches: 52  
Query Match: 13.06% Indels: 26  
DB: 14 Gaps: 6

US-10-087-573-2 (1-141) x US-10-160-758-10 (1-10531)

Qy 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22

```

; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannon, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY OF
; TITLES OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 667
; LENGTH: 1659
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-667

Alignment Scores:
Pred. No.: 0.369 Length: 1659
Score: 90.00 Matches: 40
Percent Similarity: 41.54% Conservative: 14
Best Local Similarity: 30.77% Mismatches: 41
Query Match: 12.71% Indels: 35
DB: 10 Gaps: 6

US-10-087-573-2 (1-141) x US-09-796-692-667 (1-1659)
QY 8 ThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg--- 26
Db 501 ACACGGCTTGTTGGTGGCCCTCCAGGGTACACAGCGGCTCCCGAGGGGCGATCAGGCC 442
QY 27 -----GluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42
Db 441 CCATCCGGGATCTCCACACCATGACTGTGCATCTCTCGAGGCCACTGGCAATGACGTTG 382
QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58
Db 381 TCATTG-----TCCGGGACACAGGGCGATGCTAGCACAGGGGCTGTGTGGCCACAG 331
QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74
Db 330 ACCGTGGGGCATCTTCTGTCCACACGCTCCAGTCTTGCCAGGGGCGACACCGAGAGGCC 271
QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94
Db 270 CCTCCCCCG-----CTGGCCTCACAGATC 247
QY 95 ArgGluMetAsp-----LysThrIleSerGlnGlu-SerAlaArgVa 108
Db 245 AGGGCCCAAGATTATGACCTGTCACACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 102

```

QY 108 laenHisArgLeuProGluGlyHisPro 117  
Db 186 CGCACATCTTCATAGCACTGGTGGCGCT 159

RESULT 6  
US-10-040-862-667/c  
; Sequence 667, Application US/10040862  
; Publication No. US20030078396A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Retter, Marc  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy  
; TITLE OF INVENTION: Hematological Malignancies  
; FILE REFERENCE: 014058-01352035  
; CURRENT APPLICATION NUMBER: US/10/040,862  
; CURRENT FILING DATE: 2001-11-06  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: US 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: US 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: US 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: US 60/223,416  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: US 60/223,378  
; PRIOR FILING DATE: 2000-08-07  
; PRIOR APPLICATION NUMBER: US 09/796,692  
; PRIOR FILING DATE: 2001-03-01  
; NUMBER OF SEQ ID NOS: 10467  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 667  
; LENGTH: 1659  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-040-862-667

Alignment Scores:  
Pred. No.: 0.369 Length: 1659  
Score: 90.00 Matches: 40  
Percent Similarity: 41.54% Conservative: 14  
Best Local Similarity: 30.77% Mismatches: 41  
Query Match: 12.71% Indels: 35  
DB: 14 Gaps: 6

US-10-087-573-2 (1-141) x US-10-040-862-667 (1-1659)

QY 8 ThrAspPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg--- 26  
Db 501 ACACGCTTGGTGGCCCTCCAGGGTGACAGCGGGCTCCCGAGGGCGACGATCAGGCC 442

QY 27 -----GluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
Db 441 CCATCCGGGATCCACACCATGATGTCAGTCTCCGAGGCCACTGGCANTGACGTTG 382

QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58

Db 381 TCATTG-----TGGGGCACCAGGCGATGCTAGCACAGGGGCTGTGTGCCACAG 331  
QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
Db 330 ACCGTGGCGCATTTCTTCCACACGTCCTCCAGGCGGCGACACAGGAAGGCC 271  
QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
Db 270 CTTCCCGCG-----CTGGGCTCACAGATC 247  
QY 95 ArgGluMetAsn-----LysThrIleSerGlnGlu-SerAlaArgVa 108  
Db 246 AGGCCCAAAACTTAGGGTTGACAGCACAGAGCCACTGCCAGGTGTCTGTGAGACG 187  
QY 108 laenHisArgLeuProGluGlyHisPro 117  
Db 186 CGCACATCTTCATAGCACTGGTGGCGCT 159

RESULT 7  
US-09-918-995-36271/c  
; Sequence 36271, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 36271  
; LENGTH: 432  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-918-995-36271

Alignment Scores:  
Pred. No.: 0.0952 Length: 432  
Score: 88.50 Matches: 34  
Percent Similarity: 44.66% Conservative: 12  
Best Local Similarity: 33.01% Mismatches: 27  
Query Match: 12.50% Indels: 30  
DB: 11 Gaps: 5

US-10-087-573-2 (1-141) x US-09-918-995-36271 (1-432)

QY 30 LeuArgValLysSerSerGluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGly 49  
Db 428 ATGACTGTGCAGTCTCTCGAGCCACTGGCAATGACGTTGTCTTG-----TGGGG 378

QY 50 HisArgValLeuProGlyThrGlyAla-----SerAlaIleAlaAla 63  
Db 377 CACGAGCGATGCTTAGCACAGGGGCTGTGTGGCCACAGACCGTGGCGGCATTCTTGTC 318

QY 64 Thr-----ValThrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSer 81  
Db 317 ACAGTCCAGTCTTGGCCAGGGGCGACGACGAGAGCCCTCCCGG----- 270

QY 82 ThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsn----- 98  
Db 269 -----CTGGCTCACAGATCAGGCCACAAACTTAGGGTTG 234

QY 99 -----LysThrIleSerGlnGlu-SerAlaArgValAsnHisArgLeuProGluGl 115  
Db 233 ACAGCACAGAGCCACTGTCTCCAGGTGTTCTGTGAGAGCGGCACATCTTCATAGCACTGG 174

QY 115 yHisPro 117  
Db 173 TCGGCGCT 167

RESULT 8
US-09-918-995-36013/c
; Sequence 36013, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36013
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-36013
Alignment Scores:
Pred. No.: 0.0996 Length: 447
Score: 88.50 Matches: 34
Percent Similarity: 44.66% Conservative: 12
Best Local Similarity: 33.01% Mismatches: 27
Query Match: 12.50% Indels: 20
DB: 11 Gaps: 5
US-10-087-573-2 (1-141) x US-09-918-995-36013 (1-447)
Qy 30 LeuArgValLysSerSerGluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGly 49
Db 389 ATGATGTCAGTCTTCGAGGCATCGCAATGACGTTGTCTATG-----TGGGG 339
Qy 50 HisArgValLeuProGlyThrGlyAla-----SerAlaIleAlaAla 63
Db 338 CACCAGGCGATCTCTAGCACAGGGGCTGTGGCCACAGACCGTGGGGCGCATTTGTGCC 279
Qy 64 Thr-----ValThrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSer 81
Db 278 ACGCGTCAGTCTTCCAGGGGAGCAGCAGGAGGCGCCCTCCCGC----- 231
Qy 82 ThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsn----- 98
Db 230 -----CTGGCCTCACAGATCAGGGCCACAACTTAGGGTTG 195
Qy 99 -----LysThrIleSerGlnGlu-SerAlaArgValAsnHisArgLeuProGluGl 115
Db 194 TCAGCACAGACCCACTGTCAGAGTGGTCTGTGAGCGCGCACATCTTCATAGCACTGG 135
Qy 115 yHisPro 117
Db 134 TCGGCT 128
RESULT 9
US-09-918-995-8937
; Sequence 8937, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8937
; LENGTH: 486
US-10-087-573-2 (1-141) x US-09-918-995-8937 (1-486)
Qy 20 GluThrPheAspValMetArgGluAlaLeuLeuArg-----Val 32
Db 160 GATCGTTTAAAGTTATGAGGCAAGAAATGAGGAACAGAAATATCAAGATGAAGTT 219
Qy 33 LysSerSerGluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgVal 52
Db 220 AAGAAGCTTCAAAAGGCAGCTCTCTCTGAAAGCT-----TCTCCCAAAAAA 264
Qy 53 LeuProGlyThr---GlyAlaSerAlaIleAlaAlaThrValThrProLysGlyAlaSer 71
Db 265 GCACCTGGTACTAAAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 324
Qy 72 MetLysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSer 91
Db 325 GTTCAGCAAAAAGATCACCGCGCGAGTAAAGAGGCTCCAGCCCATAGGTTCTGCCCC 384
Qy 92 ArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArg 111
Db 385 AGAAGCCACAGCCCATAAAGCAGCGCTGCTCTCAAAAGCTCAGAGGGTCAAAAAGCTC 444
Qy 112 LeuProGluGlyHisProLeuGluGlyAlaArgAlaGluTyrPheArgHisLeu 129
Db 445 CAGCCCAAGAGGAC---CTGCTCCAA-----AGGCATCTG 477
RESULT 10
US-09-864-636A-89
; Sequence 89, Application US/09864636A
; Publication No. US20030104378A1
; GENERAL INFORMATION:
; APPLICANT: Third Wave Technologies
; APPLICANT: Allwal, Hatim
; APPLICANT: Bartholomay, Christian
; APPLICANT: Chehak, LuAnne
; TITLE OF INVENTION: Detection of RNA Sequences
; FILE REFERENCE: FORS-04944
; CURRENT APPLICATION NUMBER: US/09/864,636A
; CURRENT FILING DATE: 2002-10-15
; NUMBER OF SEQ ID NOS: 2640
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 89
; LENGTH: 2517
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-864-636A-89
Alignment Scores:
Pred. No.: 3.51 Length: 2517
Score: 84.00 Matches: 55
Percent Similarity: 41.48% Conservative: 18
Best Local Similarity: 31.25% Mismatches: 62
Query Match: 11.86% Indels: 42
DB: 11 Gaps: 9

```

US-10-087-573-2 (1-141) x US-09-864-636A-89 (1-2517)

QY 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20
    : : : : :
Db 1141 CTGGACCTTCGAACACACACCCCGAGGGGCTGGCG-----CGGGCTACGGGGGGGAG 1194
    : : : : :

QY 21 ---ThrPheAspValMetArgGluAlaLeuLeu----- 30
    : : : : :
Db 1195 TGGACGAGAGACCGCCGCCACCGGGCCCTCTCTCGAGAGGCTCCATCGGAACCTCCCTT 1254
    : : : : :

QY 31 ---ArgValIysSerSerGluArgLeuAlaMetLeu----- 41
    : : : : :
Db 1255 AAGCGCTCTCAGGGGGAGGAGAGCTCCTTTGGCTCTACCACGAGGTGGAAAAAGCCCCCTC 1314
    : : : : :

QY 42 ---ArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSer--- 59
    : : : : :
Db 1315 TCCGGGCTCTGCCCATATGGA--GGCCACGGGGGTCGGCTGGACCTGGCCTATCTCAG 1373
    : : : : :

QY 60 -----AlaIleAlaAlaThrValThrProLysGlyAlaSerMetLysLeu 74
    : : : : :
Db 1374 GGCCTTGTCCCTCGAGGCTTCGCGAGGAGATCGCCCGCTCGAGCGCCAGGT-----CTT 1427
    : : : : :

QY 75 LysProProArgPro-----GlnSerThrLysSerProGlu 86
    : : : : :
Db 1428 CCGCTGGCGGGCCACCCCTTCAACCTCAACTCCCGGACCAGCTGGAAGGGTCTCTT 1487
    : : : : :

QY 87 LeuArgGluLeuSerArgLysIleArgGlu--MetAsnLysThrIleSerGlnGlu---Se 105
    : : : : :
Db 1488 TGACGAGCTAGGGCTTCCCGCCATCGGCCAAGACGGAGAAGACCGGAAGCGCTCCACCAG 1547
    : : : : :

QY 105 rAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuLysArgAlaGluTy 125
    : : : : :
Db 1548 CGCGCCGCTCTCGAGGCCCTTCGCGAGGCCACCCCATCGTGGAGGAAGATCCTCGAGTA 1607
    : : : : :

QY 125 rPheArgHisLeuArgSerLeuIysSerGlnGlyValAsnArgLeu 140
    : : : : :
Db 1608 C---CGGGAGCTCACCAGGTGAGAGACACCTACATTGACCCCTTG 1650
    : : : : :

```

```

RESULT 11
US-09-758-282-120
; Sequence 120, Application US/09758282
; Publication No. US20030134349A1
; GENERAL INFORMATION:
; APPLICANT: Ma, Wu-Po
; APPLICANT: Lvamichev, Victor I.
; APPLICANT: Kaiser, Michael W.
; APPLICANT: Lyamicheva, Natalie E.
; APPLICANT: Allawi, Hatim T.
; APPLICANT: Schaefer, James J.
; APPLICANT: Neri, Bruce P.
; TITLE OF INVENTION: Improved Enzymes for the Detection of Specific Nucleic
; FILE REFERENCE: FORS-04323
; CURRENT APPLICATION NUMBER: US/09/758,282
; CURRENT FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 280
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 120
; LENGTH: 2517
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-758-282-120

Alignment Scores:
Pred. No.: 3.51
Score: 84.00
Percent Similarity: 41.48%
Best Local Similarity: 31.25%
Query Match: 11.86%
DB: 12
Matches: 55
Conservative: 18
Mismatchches: 62
Indels: 42
Gaps: 9

```

; SEQ ID NO 89  
; LENGTH: 2517  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-10-084-839-89  
Alignment Scores:  
Pred. No.: 3.51 Length: 2517  
Score: 84.00 Matches: 55  
Percent Similarity: 41.48% Conservative: 18  
Best Local Similarity: 31.25% Mismatches: 62  
Query Match: 11.86% Indels: 42  
DB: 12 Gaps: 9  
US-10-087-573-2 (1-141) x US-10-084-839-89 (1-2517)  
Qy 1 MetGluSerThrSerThrThrThrAspPheValAlaGluAsnArgProThrPheGlyGlu 20  
Db 1141 CTGACCTCTTCGACACACCCCGAGGGGTGGCG-----CGGCGCTACGGGGGGAG 1194  
Qy 21 --ThrPheAspValMetArgGluAlaLeu----- 30  
Db 1195 TGGACGGAGGACGGCGCCCGCCCTCTCTCGGAGAGGCTCCATCGGAACCTCCTT 1254  
Qy 31 --ArgValLysSerSerGluArgLeuAlaMetLeu----- 41  
Db 1255 AAGCGCTCGAGGGGAGGAGAACTCTTTGGCTTACCACGAGGTGGAAAGCCCTC 1314  
Qy 42 --ArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSer--- 59  
Db 1315 TCCCGGTCTTGGCCCATATGGA--GGCACGGGGTGGCTGTGACGTGCCTATCTCAG 1373  
Qy 60 -----AlaIleAlaAlaThrValThrProLysGlyAlaSerMetLysLeu 74  
Db 1374 GGCCTTGTCTGAGGCTTGGCGAGGAGATCGCCCGCTCGAGGCGAGGT-----CTT 1427  
Qy 75 LysProProArgPro-----GlnSerThrLysSerProGlu 86  
Db 1428 CCGCTGCGCGGCCACCCCTTCAACTCCCGGAGCAGCGGCAAGCGCTCCACAG 1547  
Qy 87 LeuArgGluLeuSerArgLysIleArgGlu--MetAsnLysThrIleSerGlnGlu---Se 105  
Db 1488 TGACGAGTAGGGCTTCGCGCCATCGCAAGACGAGAGACCGGCAAGCGCTCCACAG 1547  
Qy 105 xAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGluTy 125  
Db 1548 CGCGCGCTCTGTGAGGCCCTTCGCGAGGCCACCCCATCGTGAGAGATCTCTGCAGTA 1607  
Qy 125 rPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140  
Db 1608 C---CGGAGCTCACCAAGCTGAAGACACCTACATTGACCCCTTG 1650  
RESULT 13  
US-10-029-386-5420  
; Sequence 5420, Application US/10029386  
; Publication No. US2003019704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G  
; FILE REFERENCE: AEMICA-X-2  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 34288  
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 5420  
; LENGTH: 505  
; TYPE: DNA  
; ORGANISM: Homo sapiens

; FEATURE:  
; OTHER INFORMATION: MAP TO AC012331.9  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.3  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.2  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.1  
; OTHER INFORMATION: EST HUMAN HIT: B1906112.1, EVALUE 3.00e-31  
; OTHER INFORMATION: NT HIT: 273645.1, EVALUE 0.00e+00  
; OTHER INFORMATION: SWISSPROT HIT: P01714, EVALUE 6.00e-30  
US-10-029-386-5420  
Alignment Scores:  
Pred. No.: 0.64 Length: 505  
Score: 82.50 Matches: 31  
Percent Similarity: 43.48% Conservative: 9  
Best Local Similarity: 33.70% Mismatches: 41  
Query Match: 11.65% Indels: 11  
DB: 12 Gaps: 4  
US-10-087-573-2 (1-141) x US-10-029-386-5420 (1-505)  
Qy 35 SerGluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuPro 54  
Db 224 AGCCAGAGAAATCGCTCAGGATCCTTGGGGCGGTCTACTGTATCATAGA---TGACCA 280  
Qy 55 GlyThrGlyAlaSer---AlaIleAlaAlaThrValThrProLysGlyAlaSerMetLys 73  
Db 281 GCACGGGGCGCTGCTGCTGTGTACCGAGTGTTCATAAGAGCCTTCCATGCTGT 340  
Qy 74 LeuLys-----ProProArgProGlnSerThrLysSerProGlu 86  
Db 341 CTCCTGGCAGGTGATCTCGGCCATTTGTCCCAAGGCCACAGACACTGCGACGACCTGAG 400  
Qy 87 LeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAla 106  
Db 401 TTGGCCAG-----AGGAAATACGGAGCCCAAGAGGAGGAGGAGGTGAGTCTG 454  
Qy 107 ArgValAsnHisArgLeuProGluGlyHisProLeu 118  
Db 455 AGGATGAGGGCGCTCATTCACAGAACCCACACCCCTG 490  
RESULT 14  
US-10-156-761-7288  
; Sequence 7288, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: HORIKAWA, JUN  
; APPLICANT: ISHIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 7288  
; LENGTH: 747  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(747)  
US-10-156-761-7288  
Alignment Scores:

Pred. No.: 1.08 Length: 747  
 Score: 82.50 Matches: 30  
 Percent Similarity: 40.66% Conservative: 7  
 Best Local Similarity: 32.97% Mismatches: 34  
 Query Match: 11.65% Indels: 20  
 DB: 14 Gaps: 3  
 US-10-087-573-2 (1-141) x US-10-156-761-7288 (1-747)  
 QY 58 AlaSerAlaIleAlaThrValThrProLysGlyAlaSerMetLysLeuLysProPro 77  
 Db 473 GCATCGCTCGCGCGCTGTGTGGCGCGCTGCCGAAAGCGCGCGGATGCTGTTTCACT 532  
 QY 78 ArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMet 97  
 Db 533 CGGCCAAGATCAACCACTCAAGCTGCTGCCGAGGGCGCGCGGAGAGCCGCG 592  
 QY 98 AsnLysThr-----IleSerGlnGluSer-AlaAr 107  
 Db 593 TCCTCGACATGTTGACCCAGATGACGAGGGCTTCGGTGGCTGCACCTCGCGGAG 652  
 QY 107 gValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGluTyrPheAr 127  
 Db 653 AGTGCGCCACCGCTGCCGAGGGATCCC-----GCTGGTCTCCATCACCG 700  
 QY 127 gHisLeuArgSerLeuLysSerGlnGlyVal 137  
 Db 701 GCAT-----GAACAAGGAGTG 716  
 RESULT 15  
 US-10-156-761-1  
 ; Sequence 1, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; PRIOR FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 1  
 ; LENGTH: 9025608  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces avermitilis  
 ; NAME/KEY: misc feature  
 ; LOCATION: (4187715)  
 ; OTHER INFORMATION: a, t, c, g, other or unknown  
 US-10-156-761-1  
 Alignment Scores:  
 Pred. No.: 2.63e+05 Length: 9025608  
 Score: 82.50 Matches: 30  
 Percent Similarity: 40.66% Conservative: 7  
 Best Local Similarity: 32.97% Mismatches: 34  
 Query Match: 11.65% Indels: 20  
 DB: 14 Gaps: 3  
 US-10-087-573-2 (1-141) x US-10-156-761-1 (1-9025608)  
 QY 58 AlaSerAlaIleAlaThrValThrProLysGlyAlaSerMetLysLeuLysProPro 77  
 Db 8718299 GCATCGCTCGCGCGCTGTGTGGCGCGCTGCCGAAAGCGCGCGGATGCTGTTTCACT 8718358

QY 78 ArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMet 97  
 Db 8718359 CGGCCAAGATCAACCACTCAAGCTGCTGCCGAGGGCGCGCGGATGCTGTTTCACT 8718418  
 QY 98 AsnLysThr-----IleSerGlnGluSer-AlaAr 107  
 Db 8718419 TCCTCGACATGTTGACCCAGATGACGAGGGCTTCGGTGGCTGCACCTCCCGGAG 8718478  
 QY 107 gValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGluTyrPheAr 127  
 Db 8718479 AGTGCGCCACCGCTGCCGAGGGATCCC-----GCTGGTCTCCATCACCG 8718526  
 QY 127 gHisLeuArgSerLeuLysSerGlnGlyVal 137  
 Db 8718527 GCAT-----GAACAAGGAGTG 8718542  
 RESULT 16  
 US-10-156-761-3722  
 ; Sequence 3722, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; PRIOR FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 3722  
 ; LENGTH: 786  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces avermitilis  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(786)  
 US-10-156-761-3722  
 Alignment Scores:  
 Pred. No.: 1.33 Length: 786  
 Score: 82.00 Matches: 47  
 Percent Similarity: 35.98% Conservative: 21  
 Best Local Similarity: 24.87% Mismatches: 58  
 Query Match: 11.58% Indels: 64  
 DB: 14 Gaps: 8  
 US-10-087-573-2 (1-141) x US-10-156-761-3722 (1-786)  
 QY 4 ThrSerThrThrThrAsnPheValAlaGluAsnArgProThr-----Phe 18  
 Db 4 ACCTCACTCCGCGACGTCGCTTGGCGACCTGGACCCCGAGCGATCGTGGCGTGTTC 63  
 QY 19 GlyGluThrPheAspValMetArgGluAlaLeu-----LeuArgValLysSerSerGlu 36  
 Db 64 CGGCAGATCGCGCACCGACGTCGCGAGGCCATCGACCGTGGCGGATTCAGAGGGCGGAA 123  
 QY 37 ArgLeu-----AlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgVal 52  
 Db 124 AAGCTCCCTCGGAGGCTGAGCTCGT-CGAGGACTACGGAGTGTCCCGGATGACCGTCCG 182  
 QY 53 -----LeuProGlyThrGlyAlaSerAlaIleAlaIle----- 63  
 Db 183 GAACCTCTTCTCCATCTCCAGGCGGAGGGCTCGTCCATCGCCGAGCACCGAAGGCGT 242



```

QY 65 ValThrProLysGlyAlaSerMetLysLeuLysPro-----ProArg 78
   |||||
Db 6000095 GTGATCCCGTACTCCGCGAGCTTGCCCTTGCGCGCGTCCGGGGCGGTGAGGACCAAGCGGG 6000036
QY 79 ProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsn 98
   |||||
Db 6000035 CCCTCTCCGTCCGAGCCACCGAGTGCTCCAGTCCGAGGACCAGTCCCGTCGTGGTGTG 5999976
QY 99 LysThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeu 118
   |||||
Db 5999975 ATGACGGTCCAGTCGCTCTCGAGCAC-CTCGGTCTTTCGGTGTGTCGAGGGACACCATCGG 5999917
QY 119 LeuGluLysArgAlaGluTyPheArgHisIleArgSerLeuLysSerGlnGlyValAsn 138
   |||||
Db 5999916 CTCGATCCCGAGGAGAGCCGGGACCACGTTTCGACCCTTGCCGCGCGCGGTCCAC 5999857

RESULT 18
US-08-319-974A-1
; Sequence 1, Application US/08319974A
; Publication No. US20030104576A1
; GENERAL INFORMATION:
; APPLICANT: Nyce, Jonathan W.
; TITLE OF INVENTION: Attenuated Viruses and Method of Making
; TITLE OF INVENTION: the Same
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Kenneth D. Sibley
; STREET: Post Office Box 34009
; CITY: Charlotte
; STATE: No. US20030104576A1th Carolina
; COUNTRY: USA
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/319,974A
; FILING DATE: 07-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5218-27
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-881-3140
; TELEFAX: 919-881-3175
; TELEX: 575102
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9718 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-319-974A-1

Alignment Scores:
Pred. No.: 42.6 Length: 9718
Score: 81.50 Matches: 34
Percent Similarity: 37.50% Conservative: 11
Best Local Similarity: 28.33% Mismatches: 37
Query Match: 11.51% Indels: 38
DB: 8 Gaps: 6

US-10-087-573-2 (1-141) x US-08-319-974A-1 (1-9718)
QY 22 PheAspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeu 41
Db 5373 TTCGACTGTTTTCGGACTCCGGATACGC-----MetCysGlyHisArgVal 5402
QY 42 ArgAlaLeuAlaGly-----MetCysGlyHisArgVal 52

```

```

; Sequence 1, Application US/09833102
; Patent No. US2002006652A1
; GENERAL INFORMATION:
; APPLICANT: Danielson, Steffen
; APPLICANT: Schneider, Palle
; TITLE OF INVENTION: Nucleic acids encoding polypeptides having haloperoxidase activity
; FILE REFERENCE: 10172.200-US
; CURRENT APPLICATION NUMBER: US/09/833,102
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 1
; LENGTH: 1797
; TYPE: DNA
; ORGANISM: Phaeotrichoconis crotalariae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1797)
US-09-833-102-1

Alignment Scores:
Pred. No.: 5,26 Length: 1797
Score: 81.00 Matches: 27
Percent Similarity: 38.64% Conservatives: 7
Best Local Similarity: 30.68% Mismatches: 27
Query Match: 11.44% Indels: 27
DB: Gaps: 3

US-10-087-573-2 (1-141) x US-09-833-102-1 (1-1797)
Qy 54 ProGlyThrGlyAlaSerAlaIleAlaThrValThrProLysGlyAlaSerMetLys 73
Db 1073 CCGCGGTACGGATGACGGCGGCCAGACACGGCGCCCTTCTGGCTCACCTCGCGG 1132
Qy 74 LeuLysPro-----ProArgProGlnSerThrLysSerProGluLeuArgGlu 89
Db 1133 CTCCTGCCACAAACACAAACGACATCCCTTCAAGCCCCCTTCCCGCTACCCGTCGG 1192
Qy 90 LeuSerArgLysIle-ArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValas 109
Db 1193 GCCACGGACCTTTGGCGG-----
Qy 109 nHisArgLeuProGluGlyHisProLeuLeuGluLys----- 121
Db 1212 -CGCCGTCTTCCAAATGGTGGCGCGCTACTACAACGGCGCGCTCGGCACGTGGAGGACA 1270
Qy 122 -ArgAlaGluTyrPheArgHis 128
Db 1271 ACGAGCCCGACACATCGCCAT 1292

RESULT 20
US-09-832-441-1
; Sequence 1, Application US/09832441
; Patent No. US2002009434A1
; GENERAL INFORMATION:
; APPLICANT: Danielson, Steffen
; APPLICANT: Schneider, Palle
; TITLE OF INVENTION: Polypeptides having haloperoxidase activity
; FILE REFERENCE: 10019.200-US
; CURRENT APPLICATION NUMBER: US/09/832,441
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 1
; LENGTH: 1797
; TYPE: DNA
; ORGANISM: Phaeotrichoconis crotalariae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1797)
US-09-832-441-1

Alignment Scores:
Pred. No.: 5,26 Length: 1797
Score: 81.00 Matches: 27
Percent Similarity: 38.64% Conservatives: 7
Best Local Similarity: 30.68% Mismatches: 27
Query Match: 11.44% Indels: 27
DB: Gaps: 3

US-10-087-573-2 (1-141) x US-09-832-441-1 (1-1797)
Qy 54 ProGlyThrGlyAlaSerAlaIleAlaThrValThrProLysGlyAlaSerMetLys 73
Db 1073 CCGCGGTACGGATGACGGCGGCCAGACACGGCGCCCTTCTGGCTCACCTCGCGG 1132
Qy 74 LeuLysPro-----ProArgProGlnSerThrLysSerProGluLeuArgGlu 89
Db 1133 CTCCTGCCACAAACACAAACGACATCCCTTCAAGCCCCCTTCCCGCTACCCGTCGG 1192
Qy 90 LeuSerArgLysIle-ArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValas 109
Db 1193 GCCACGGACCTTTGGCGG-----
Qy 109 nHisArgLeuProGluGlyHisProLeuLeuGluLys----- 121
Db 1212 -CGCCGTCTTCCAAATGGTGGCGCGCTACTACAACGGCGCGCTCGGCACGTGGAGGACA 1270
Qy 122 -ArgAlaGluTyrPheArgHis 128
Db 1271 ACGAGCCCGACACATCGCCAT 1292

RESULT 21
US-10-156-761-2584/c
; Sequence 2584, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAVOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2584
; LENGTH: 1902
```

```

; Sequence 1, Application US/098332441
; Patent No. US2002009434A1
; GENERAL INFORMATION:
; APPLICANT: Danielson, Steffen
; APPLICANT: Schneider, Palle
; TITLE OF INVENTION: Polypeptides having haloperoxidase activity
; FILE REFERENCE: 10019.200-US
; CURRENT APPLICATION NUMBER: US/09/832,441
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 1
; LENGTH: 1797
; TYPE: DNA
; ORGANISM: Phaeotrichoconis crotalariae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1797)
US-09-832-441-1

Alignment Scores:
Pred. No.: 5,26 Length: 1797
Score: 81.00 Matches: 27
Percent Similarity: 38.64% Conservatives: 7
Best Local Similarity: 30.68% Mismatches: 27
Query Match: 11.44% Indels: 27
DB: Gaps: 3

US-10-087-573-2 (1-141) x US-09-832-441-1 (1-1797)
Qy 54 ProGlyThrGlyAlaSerAlaIleAlaThrValThrProLysGlyAlaSerMetLys 73
Db 1073 CCGCGGTACGGATGACGGCGGCCAGACACGGCGCCCTTCTGGCTCACCTCGCGG 1132
Qy 74 LeuLysPro-----ProArgProGlnSerThrLysSerProGluLeuArgGlu 89
Db 1133 CTCCTGCCACAAACACAAACGACATCCCTTCAAGCCCCCTTCCCGCTACCCGTCGG 1192
Qy 90 LeuSerArgLysIle-ArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValas 109
Db 1193 GCCACGGACCTTTGGCGG-----
Qy 109 nHisArgLeuProGluGlyHisProLeuLeuGluLys----- 121
Db 1212 -CGCCGTCTTCCAAATGGTGGCGCGCTACTACAACGGCGCGCTCGGCACGTGGAGGACA 1270
Qy 122 -ArgAlaGluTyrPheArgHis 128
Db 1271 ACGAGCCCGACACATCGCCAT 1292

RESULT 21
US-10-156-761-2584/c
; Sequence 2584, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAVOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2584
; LENGTH: 1902
```

LENGTH: 2517  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(1902)  
US-10-156-761-2584

Alignment Scores:  
Pred. No.: 5.67 Length: 1902  
Score: 81.00 Matches: 32  
Percent Similarity: 44.33% Conservative: 11  
Best Local Similarity: 32.9% Mismatches: 29  
Query Match: 11.44% Indels: 25  
DB: 14 Gaps: 5

US-10-087-573-2 (1-141) x US-10-156-761-2584 (1-1902)

Qy 39 AlaMetLeuArgAlaLeuAlaGlyMet-----CysGlyHisArgValLeuPro--- 54  
Db 1022 GCCCGCGTGGCGCGCGCGCGGTTTGGCGCGAGCGCCAGCGAGGTGAACCGGCC 963  
Qy 55 GlyThrGlyAlaSerAlaIleAlaIleAlaThrValThrPro-LysGlyAlaSerMetLysLe 74  
Db 962 GGGACCGGAACAGCGCGATC-----CCGGCGAGGAGCGCGTCTGCCGGT 918  
Qy 74 ulysProProArgPro-----GlnSerThrLysSerPro----- 85  
Db 917 AGCAGCGCTCGCCCGCGCGCGCGAGTACCGCGAGCCCGGATGTGCGCGCGCGCCCTC 858  
Qy 86 -----GluLeuArgGluLeuSerArgLysIleArgGluMetAsnLy 99  
Db 857 TTGCGCCACACCGAGTCGCGCGCGCGTCCGCTCCAGCGCGCGGTCAATGTCACGAGCTCGGG 798  
Qy 99 sThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGly 115  
Db 797 CAGGACGAGTCGAGGCGCAGCAGATGCGCGCGCGGTCTCGAGCGGG 749

#### RESULT 22

US-10-084-839-2690  
Sequence 2690, Application US/10084839  
Publication No. US20030186238A1

#### GENERAL INFORMATION:

APPLICANT: Third Wave Technologies  
APPLICANT: Allawi, Hatim  
APPLICANT: Argue, Brad T.  
APPLICANT: Bartholomay, Christian T.  
APPLICANT: Chehak, LuAnne  
APPLICANT: Curtis, Michelle L.  
APPLICANT: Eis, Peggy S.  
APPLICANT: Hall, Jeff G.  
APPLICANT: Ip, Hon S.  
APPLICANT: Ji, Lin  
APPLICANT: Kaiser, Michael  
APPLICANT: Kwiatkowski, Jr., Robert W.  
APPLICANT: Lukowiak, Andrew A.  
APPLICANT: Lymaichev, Victor  
APPLICANT: Lymaicheva, Natalie E.  
APPLICANT: Ma, Wufo  
APPLICANT: Neri, Bruce P.  
APPLICANT: Olson, Sarah M.  
APPLICANT: Olson-Munoz, Marilyn C.  
APPLICANT: Schaefer, James J.  
APPLICANT: Skrzypczynski, Zbigniew  
APPLICANT: Takova, Tsetska Y.  
APPLICANT: Thompson, Lisa C.  
APPLICANT: Vedvik, Kevin L.  
TITLE OF INVENTION: RNA Detection Assays  
FILE REFERENCE: FORS-06666  
CURRENT APPLICATION NUMBER: US/10/084,839  
CURRENT FILING DATE: 2002-02-26  
NUMBER OF SEQ ID NOS: 4004  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 2690

LENGTH: 2517  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Synthetic  
US-10-084-839-2690

Alignment Scores:  
Pred. No.: 8.21 Length: 2517  
Score: 81.00 Matches: 55  
Percent Similarity: 41.04% Conservative: 16  
Best Local Similarity: 31.79% Mismatches: 61  
Query Match: 11.44% Indels: 42  
DB: 12 Gaps: 9

US-10-087-573-2 (1-141) x US-10-084-839-2690 (1-2517)

Qy 4 ThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu---ThrPhe 22  
Db 1150 TCGAACACACACCCCGAGGGGTGGCG-----CGGCGCTACGGGGGGAGTGACCGGAG 1203  
Qy 23 AspValMetArgGluAlaLeuLeu-----ArgVal 32  
Db 1204 GACGCGCCACCGGGCCCTCTCGGAGAGGCTCCATCGGAACCTCTTAAGGGCCTC 1263  
Qy 33 LysSerSerGluArgLeuAlaMetLeu-----ArgAla 43  
Db 1264 GAGGGGAGGAGAGCTCTTTCGCTCTACCGAGGTGGAAGGCCCTCTTCCCGGTC 1323  
Qy 44 LeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSer----- 59  
Db 1324 CTGGCCCATATGA-GGCCACGCGGGTGGCGCTGAGCGTGGCTATCTCAGGGCCTTGTC 1382  
Qy 60 -----AlaIleAlaIleThrValThrProLysGlyAlaSerMetLysLysProPro 77  
Db 1383 CTGGAGCTTGGCGAGGATCGCCCGCTCGAGGCGGAGGT-----CTTCCGCTGGC 1436  
Qy 78 ArgPro-----GlnSerThrLysSerProGluLeuArgGlu 89  
Db 1437 CGGCCACCCCTTCAACTCCCGGACAGCTGGAAGGGTCTCTTTGACGAGCT 1496  
Qy 90 LeuSerArgLysIleArgGlu-MetAsnLysThrIleSerGlnGlu---SerAlaArgVa 108  
Db 1497 AGGCTTCCCGCATCGCAAGACGGAAGACCGCAAGCGCTCCACACAGCGCGCGT 1556  
Qy 108 lAsnHisArgLeuProGlyHisProLeuLeuGluLysArgAlaGluTyPheArgHi 128  
Db 1557 CTGGAGGCGCTCGGAGGCCCGCCCATCTGTGGAGAGATCTCTGCAGTAC---CGGA 1613  
Qy 128 sLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140  
Db 1614 GCTCAGCAAGCTGAGAGCACCTACATTGACCCCTTG 1650

#### RESULT 23

US-09-969-708-93/c  
Sequence 93, Application US/09969708  
Patent No. US20020102532A1

#### GENERAL INFORMATION:

APPLICANT: Augustus, Meena  
TITLE OF INVENTION: Cancer  
FILE OF INVENTION: Sets  
FILE REFERENCE: 589290-70  
CURRENT APPLICATION NUMBER: US/09/969,708  
CURRENT FILING DATE: 2001-10-03  
PRIOR APPLICATION NUMBER: US/60/237,606  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: US/60/237,608  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: US/60/237,425  
PRIOR FILING DATE: 2000-10-03  
NUMBER OF SEQ ID NOS: 658  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 93

US-09-969-708-93/c  
Sequence 93, Application US/09969708  
Patent No. US20020102532A1  
GENERAL INFORMATION:  
APPLICANT: Augustus, Meena  
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signal  
FILE OF INVENTION: Sets  
FILE REFERENCE: 589290-70  
CURRENT APPLICATION NUMBER: US/09/969,708  
CURRENT FILING DATE: 2001-10-03  
PRIOR APPLICATION NUMBER: US/60/237,606  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: US/60/237,608  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: US/60/237,425  
PRIOR FILING DATE: 2000-10-03  
NUMBER OF SEQ ID NOS: 658  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 93

```

; LENGTH: 3530
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-969-708-93

Alignment Scores:
Pred. No.: 12.8 Length: 3530
Score: 81.00 Matches: 30
Percent Similarity: 38.04% Conservative: 5
Best Local Similarity: 32.61% Mismatches: 32
Query Match: 11.44% Indels: 25
DB: 10 Gaps: 2

US-10-087-573-2 (1-141) x US-09-969-708-93 (1-3530)

QY 31 ArgValLysSerSerGluArgLeuAlaMet-LeuArgAlaLeuAlaGlyMetCysGlyHi 50
Db 595 CGCACGCTCAGCTCTCCAAAGTTGGCTTCCCGCCGCGGCGGCTCTCTCGCGGAGTATGCGACATTA 536

QY 50 sArgValLeu-----ProGlyThrGlyAlaSerAlaAlaAlaThrVa 65
Db 535 CCGAGTTGTATCATACACGTGGCGGCGGCGGCGGCGGCTCTCTCGCGGAGTATGCGACATTA 476

QY 50 sArgValLeu-----ProGlyThrGlyAlaSerAlaAlaAlaThrVa 65
Db 535 CCGAGTTGTATCATACACGTGGCGGCGGCGGCGGCGGCTCTCTCGCGGAGTATGCGACATTA 476

QY 65 lThrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSerThrLysSerPr 85
Db 475 GCGCGCCGCTCGGAGCCCGCGGCGGCGGCGGCGGCGGCTCTCTCGCGGAGTATGCGACATTA 416

QY 85 oGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSe 105
Db 415 CCAG----- 412

QY 105 rAlaArgValAsnHisArgLeuProGluGlyHis 116
Db 411 -AGACGGCTGGCGGCGGCGGCGGCGGCGGCTCTCTCGCGGAGTATGCGACATTA 379

RESULT 24
US-09-969-708-622/c
; Sequence 622, Application US/09969708
; Patent No. US20020102532A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE REFERENCE: 689290-70
; CURRENT APPLICATION NUMBER: US/09/969,708
; PRIOR FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: US/60/237,606
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,608
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,425
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 658
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 622
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-969-708-622

Alignment Scores:
Pred. No.: 12.8 Length: 3530
Score: 81.00 Matches: 30
Percent Similarity: 38.04% Conservative: 5
Best Local Similarity: 32.61% Mismatches: 32
Query Match: 11.44% Indels: 25
DB: 10 Gaps: 2

US-10-087-573-2 (1-141) x US-09-969-708-622 (1-3530)

QY 31 ArgValLysSerSerGluArgLeuAlaMet-LeuArgAlaLeuAlaGlyMetCysGlyHi 50
Db 595 CGCACGCTCAGCTCTCCAAAGTTGGCTTCCCGCCGCGGCGGCTCTCTCGCGGAGTATGCGACATTA 536

```

```

QY 50 sArgValLeu-----ProGlyThrGlyAlaSerAlaAlaAlaThrVa 65
Db 535 CCGAGTTGTATCATACACGTGGCGGCGGCGGCGGCGGCTCTCTCGCGGAGTATGCGACATTA 476

QY 65 lThrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSerThrLysSerPr 85
Db 475 GCGCGCCGCTCGGAGCCCGCGGCGGCGGCGGCGGCTCTCTCGCGGAGTATGCGACATTA 416

QY 85 oGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSe 105
Db 415 CCAG----- 412

QY 105 rAlaArgValAsnHisArgLeuProGluGlyHis 116
Db 411 -AGACGGCTGGCGGCGGCGGCGGCGGCGGCTCTCTCGCGGAGTATGCGACATTA 379

RESULT 25
US-09-880-107-3771/c
; Sequence 3771, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3771
; LENGTH: 3530
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 X67247
US-09-880-107-3771

Alignment Scores:
Pred. No.: 12.8 Length: 3530
Score: 81.00 Matches: 30
Percent Similarity: 38.04% Conservative: 5
Best Local Similarity: 32.61% Mismatches: 32
Query Match: 11.44% Indels: 25
DB: 10 Gaps: 2

US-10-087-573-2 (1-141) x US-09-880-107-3771 (1-3530)

QY 31 ArgValLysSerSerGluArgLeuAlaMet-LeuArgAlaLeuAlaGlyMetCysGlyHi 50
Db 595 CGCACGCTCAGCTCTCCAAAGTTGGCTTCCCGCCGCGGCGGCTCTCTCGCGGAGTATGCGACATTA 536

QY 50 sArgValLeu-----ProGlyThrGlyAlaSerAlaAlaAlaThrVa 65
Db 535 CCGAGTTGTATCATACACGTGGCGGCGGCGGCGGCGGCTCTCTCGCGGAGTATGCGACATTA 476

QY 65 lThrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSerThrLysSerPr 85
Db 475 GCGCGCCGCTCGGAGCCCGCGGCGGCGGCGGCGGCTCTCTCGCGGAGTATGCGACATTA 416

QY 85 oGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSe 105
Db 415 CCAG----- 412

QY 105 rAlaArgValAsnHisArgLeuProGluGlyHis 116
Db 411 -AGACGGCTGGCGGCGGCGGCGGCGGCGGCTCTCTCGCGGAGTATGCGACATTA 379

```

RESULT 26  
US-10-156-761-1190/c  
; Sequence 1190, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: HORIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10156761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1190  
; LENGTH: 993  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(993)  
US-10-156-761-1190  
Alignment Scores:  
Pred. No.: 2.76 Length: 993  
Score: 80.50 Matches: 30  
Percent Similarity: 48.84% Conservative: 12  
Best Local Similarity: 34.88% Mismatches: 41  
Query Match: 11.37% Indels: 4  
DB: 14 Gaps: 1  
US-10-087-573-2 (1-141) x US-10-156-761-1190 (1-993)  
QY 30 LeuArgValLysSerSerGluAtcLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGly 49  
Db 783 CTTCGCCGCTATGCTCGACCGCCAGCGAGCGGTTTCAGGCCGCGCTCGCGGC 724  
QY 50 HisArgValLeuProGlyThrGly---AlaSerAlaIleAlaAlaThrVal--ThrProL 68  
Db 723 GAACGCGGGTCCGGGGACCGCGATCCCGATGGCTTCCCGCGGCTTGACACCG 664  
QY 58 ySGlyAlaSerMetLysLeuLysProArgProGlnSerThrLysSerProGluLeuA 88  
Db 663 CAGGGACTTCTCG-AGATTCTACACCGAGGCTGTCCAGCACCGAGTCTAGCCGGTCA 605  
QY 88 rgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgV 108  
Db 604 GGAGCTGCTGAAGTCTGCTGGTGCGGTATCGATCACCGTGTCCGCGCGAGCGCGCA 545  
QY 108 alAsnHisArgLeu 112  
Db 544 CGAAGTCCGCGTTG 531  
RESULT 27  
US-09-864-636A-96  
; Sequence 96, Application US/09864636A  
; Publication No. US20030104378A1  
; GENERAL INFORMATION:  
; APPLICANT: Third Wave Technologies  
; APPLICANT: Allwai, Hatim  
; APPLICANT: Bartholomay, Christian  
; APPLICANT: Chehak, LuAnne  
; TITLE OF INVENTION: Detection of RNA Sequences  
; FILE REFERENCE: FORS-04944  
; CURRENT APPLICATION NUMBER: US/09/864,636A

; CURRENT FILING DATE: 2002-10-15  
; NUMBER OF SEQ ID NOS: 2640  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 96  
; LENGTH: 2517  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-09-864-636A-96  
Alignment Scores:  
Pred. No.: 10.9 Length: 2517  
Score: 80.00 Matches: 55  
Percent Similarity: 42.13% Conservative: 20  
Best Local Similarity: 30.90% Mismatches: 58  
Query Match: 11.30% Indels: 46  
DB: 11 Gaps: 10  
US-10-087-573-2 (1-141) x US-09-864-636A-96 (1-2517)  
QY 1 MetGluSerThrSerThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20  
Db 1141 CTGGACCCCTTCGAACACACACCCCGAGGGGTGGCG-----CGCGCTACGGGGGAG 1194  
QY 21 ---ThrPheAspValMetArgGluAlaLeu----- 30  
Db 1195 TGGACGGAGGACGCCGCCACCGGCCCTCTCTCGAGAGGCTCCATCGGAACCTCCTT 1254  
QY 31 ---ArgValLysSerSerGluArgLeuAlaMetLeu----- 41  
Db 1255 AAGCGCTCGAGGGGAGGAGAGCTCTTGGCTCTACACGAGGTGAAAGGCCCTC 1314  
QY 42 ---ArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla 60  
Db 1315 TCCCGGTCCTGGCCCATATGGA-GGCCACGGGGGTGGCTTGGACGTGC----- 1364  
QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProArg----- 78  
Db 1365 CTATCTCAGGGCTTGTCCCT---GGAGTGCCCGAGAGATCGCCGCTCGAGCCGA 1421  
QY 79 -----ProGln-----SerThrLysSer 84  
Db 1422 GGTCTTCCGCTGGCGGCCACCCCTTCAACTCCCGGACAGCTGGAAGGCT 1481  
QY 85 ProGluLeuArgGluLeuSerArgLysIleArgGlu-MetAsnLysThrIleSerGlnG 104  
Db 1482 CTCTTTGACGAGCTAGGCTTCCCGCATCGGACGAGACCGGAGAGACCGCAAGCGTCT 1541  
QY 104 u---SerAlaArgValAsnHisArgLeuProGlyHisProLeuLeuGlyLysArgAl 123  
Db 1542 CACCAGCGCGCGCTCTCGAGGCCCTCGCGAGGCCACCCCATCGTGAGAGATCCT 1601  
QY 123 agLutyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140  
Db 1602 GCAGTAC---CGGAGCTCACCAGCTGAAGAGACCTACGTGGACCCCTTG 1650  
RESULT 28  
US-09-758-282-141  
; Sequence 141, Application US/09758282  
; Publication No. US20030134349A1  
; GENERAL INFORMATION:  
; APPLICANT: Ma, Wu-Po  
; APPLICANT: Lyamichev, Victor I.  
; APPLICANT: Kaiser, Michael W.  
; APPLICANT: Lyamicheva, Natalie E.  
; APPLICANT: Allawi, Hatim T.  
; APPLICANT: Schaefer, James J.  
; APPLICANT: Nerri, Bruce P.  
; TITLE OF INVENTION: Improved Enzymes for the Detection of Specific Nucleic  
; TITLE OF INVENTION: Acid Sequences  
; FILE REFERENCE: FORS-04323  
; CURRENT APPLICATION NUMBER: US/09/758,282

APPLICANT: Kwiatkowski, Jr., Robert W.  
APPLICANT: Lukowiak, Andrew A.  
APPLICANT: Lyamachev, Victor  
APPLICANT: Lyamacheva, Natalie E.  
APPLICANT: Ma, WuPo  
APPLICANT: Neri, Bruce P.  
APPLICANT: Olson, Sarah M.  
APPLICANT: Olson-Munoz, Marilyn C.  
APPLICANT: Schaefer, James J.  
APPLICANT: Skrzypczynski, Zbigniew  
APPLICANT: Takova, Tsatska Y.  
APPLICANT: Thompson, Lisa C.  
APPLICANT: Thompson, Kevin L.  
APPLICANT: Vedvik, Kevin L.  
TITLE OF INVENTION: RNA Detection Assays  
FILE REFERENCE: FORS-06666  
CURRENT APPLICATION NUMBER: US/10/084,839  
CURRENT FILING DATE: 2002-02-26  
NUMBER OF SEQ ID NOS: 4004  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 96  
LENGTH: 2517  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-09-758-282-141

Alignment Scores:  
Pred. No.: 10.9 Length: 2517  
Score: 80.00 Matches: 55  
Percent Similarity: 42.13% Conservatives: 20  
Best Local Similarity: 30.90% Mismatches: 58  
Query Match: 11.30% Indels: 46  
DB: 12 Gaps: 10

US-10-087-573-2 (1-141) x US-09-758-282-141 (1-2517)

QY 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20  
Db 1141 CTGGACCCCTCGAACACACCCCGAGGGGTGGCG-----CGGCGCTACGGGGGGGAG 1194  
QY 21 ---ThrPheAspValMetArgGluAlaLeuLeu----- 30  
Db 1195 TGGACGGAGGACGGCCGCCACCGGGCCCTCTCGGAGAGGCTCCATCGGAACCTCCTT 1254  
QY 31 ---ArgValLysSerSerGluArgLeuAlaMetLeu----- 41  
Db 1255 AAGCGCCCTCGAGGGGAGGAGAGCTCTTTGGCTCTACACGAGGTGGAAAGCCCTC 1314  
QY 42 ---ArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla 60  
Db 1315 TCCCGGGTCTCGGCCCATATGGA-GGCCACGGGGGTGGCGTGCACGTGGC----- 1364  
QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArg----- 78  
Db 1365 CTATCTCAGGGCTTGTCCT---GGAGGTGGCCGAGAGATCGCCGCTCGAGGCCGA 1421  
QY 79 -----ProGln-----SerThrLysSer 84  
Db 1422 GGTCTTCGCTGGCCGCCACCCCTTCAACTCACTCCCGGGACCACTCGAAGGGT 1481  
QY 85 ProGluLeuArgGluLeuSerArgLysIleArgGlu-MetAsnLysThrIleSerGlnG 104  
Db 1482 CCTCTTCAGAGCTAGGGCTTCCGCCCATCGGAAGAGGAGAGCCGCAAGCGCTC 1541  
QY 104 u---SerAlaArgValAsnHisArgLeuProGlyHisProLeuLeuGlyLysArgAl 123  
Db 1542 CACCAGCGCGCGCTCTCGGAGGCCCTCGCGAGGCCACCCCATCGTGGAGAGATCCT 1601  
QY 123 aGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140  
Db 1602 GCAGTAC---CGGAGGCTCACCAAGCTGAGAGCACCTACGTGGACCCCTTG 1650

RESULT 29

US-10-084-839-96  
Sequence 96, Application US/10084839  
Publication No. US20030186238A1  
GENERAL INFORMATION:  
APPLICANT: Third Wave Technologies  
APPLICANT: Allawi, Hatim  
APPLICANT: Argue, Brad T.  
APPLICANT: Bartholomay, Christian T.  
APPLICANT: Chehak, LuAnne  
APPLICANT: Curtis, Michelle L.  
APPLICANT: Eis, Peggy S.  
APPLICANT: Hall, Jeff G.  
APPLICANT: JP, Hon S.  
APPLICANT: Ji, Lin  
APPLICANT: Kaiser, Michael

Alignment Scores:  
Pred. No.: 10.9 Length: 2517  
Score: 80.00 Matches: 55  
Percent Similarity: 42.13% Conservatives: 20  
Best Local Similarity: 30.90% Mismatches: 58  
Query Match: 11.30% Indels: 46  
DB: 12 Gaps: 10

US-10-087-573-2 (1-141) x US-10-084-839-96 (1-2517)

QY 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20  
Db 1141 CTGGACCCCTCGAACACACCCCGAGGGGTGGCG-----CGGCGCTACGGGGGGGAG 1194  
QY 21 ---ThrPheAspValMetArgGluAlaLeuLeu----- 30  
Db 1195 TGGACGGAGGACGGCCGCCACCGGGCCCTCTCTCGGAGAGGCTCCATCGGAACCTCCTT 1254  
QY 31 ---ArgValLysSerSerGluArgLeuAlaMetLeu----- 41  
Db 1255 AAGCGCCCTCGAGGGGAGGAGAGCTCTTTGGCTCTACACGAGGTGGAAAGCCCTC 1314  
QY 42 ---ArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla 60  
Db 1315 TCCCGGGTCTCGGCCCATATGGA-GGCCACGGGGGTGGCGTGCACGTGGC----- 1364  
QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArg----- 78  
Db 1365 CTATCTCAGGGCTTGTCCT---GGAGGTGGCCGAGAGATCGCCGCTCGAGGCCGA 1421  
QY 79 -----ProGln-----SerThrLysSer 84  
Db 1422 GGTCTTCGCTGGCCGCCACCCCTTCAACTCACTCCCGGGACCACTCGAAGGGT 1481  
QY 85 ProGluLeuArgGluLeuSerArgLysIleArgGlu-MetAsnLysThrIleSerGlnG 104  
Db 1482 CCTCTTCAGAGCTAGGGCTTCCGCCCATCGGAAGAGGAGAGCCGCAAGCGCTC 1541  
QY 104 u---SerAlaArgValAsnHisArgLeuProGlyHisProLeuLeuGlyLysArgAl 123  
Db 1542 CACCAGCGCGCGCTCTCGGAGGCCCTCGCGAGGCCACCCCATCGTGGAGAGATCCT 1601  
QY 123 aGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140  
Db 1602 GCAGTAC---CGGAGGCTCACCAAGCTGAGAGCACCTACGTGGACCCCTTG 1650

RESULT 30  
US-09-547-267-8/c  
; Sequence 8, Application US/09547267  
; Patent No. US20020147371A1  
; GENERAL INFORMATION:  
; APPLICANT: Hohmann, Hans-Peter  
; APPLICANT: Pasamontes, Luis  
; APPLICANT: Tessier, Michel  
; APPLICANT: van Loon, Adolphus  
; TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION  
; NUMBER OF SEQUENCES: 47  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: Hoffmann-La Roche Inc.  
; STREET: 340 Kingsland Street  
; CITY: Nutley  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 07110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/547,267  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/660,645  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pokras, Bruce A.  
; REGISTRATION NUMBER: 32,748  
; REFERENCE/DOCKET NUMBER: RAN 6002/170  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (201) 235-5801  
; TELEFAX: (201) 235-2363  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1149 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-09-547-267-8  
Alignment Scores:  
Pred. No.: 4.45 Length: 1149  
Score: 79.50 Matches: 35  
Percent Similarity: 42.15% Conservative: 16  
Best Local Similarity: 28.93% Mismatches: 50  
Query Match: 11.23% Indels: 20  
DB: 10 Gaps: 4  
US-10-087-573-2 (1-141) x US-09-547-267-8 (1-1149)  
QY 37 ArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArg-----51  
Db 590 CGGTATCTCGATCAGGATCGGGTGGAGTGAAGGCGACAGATAGATGAAGCGGTAC 531  
QY 52 ---ValLeuProGlyThrGlyAlaSerAlaAlaAlaAlaThrValThrProGlyGlyAla 70  
Db 530 CCGTCCATCTCGGAACGGTCCGTCATGATCATCGGGCGCTCGAGCCATCGGGGGCG 471  
QY 71 SerMetLysLeuLysProProArg-----ProGlnSerThrLysSerProgl 86  
Db 470 TCGGTCTGATCTCGAGCCCGACCAATTTCTGGAACACCCACGGTCAGGTGCGGGGTCTCG 411  
QY 86 uLeuArgGluLeuSerArgIlysIleAeAGluMetAsnLysThrIleSerGlnGluSerAl 106  
Db 410 A---CGGCACCACCGGGCGTGCATCACCAGCAGCCCTCGATCCGCGAGCGCTCGTCAGC 354  
QY 106 aArgValAsnHisArg-----LeuProGluGlyHisProLe 118

Db 353 GTCGCGCGGTATCGTCCAGCGTGCACATGCGTATTCCACCGCAGATGCACACCTGC 294  
QY 118 uLeuGluLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValas 138  
Db 293 AGCAGCCCGATCAGCGCGCGCGCTCGATGAGCCATAGCTGTCTGTCAGGCGCGGAA 234  
QY 138 n 138  
Db 233 T 233  
RESULT 31  
US-10-156-761-2702  
; Sequence 2702, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 2702  
; LENGTH: 1668  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1668)  
US-10-156-761-2702  
Alignment Scores:  
Pred. No.: 7.28 Length: 1668  
Score: 79.50 Matches: 38  
Percent Similarity: 43.51% Conservative: 19  
Best Local Similarity: 29.01% Mismatches: 59  
Query Match: 11.23% Indels: 15  
DB: 14 Gaps: 7  
US-10-087-573-2 (1-141) x US-10-156-761-2702 (1-1668)  
QY 14 AsnArgProThrPheGlyGluThrPheAspValMetArgGluAlaLeuArgValLys 33  
Db 456 CATCGGCCA-----GGAGTTGGGGCCCGTGGGGCGTGGCGTCACTCGCCCTGGGC 509  
QY 34 SerSerGluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeu 53  
Db 510 GCTGCGGACGCGTTCAGCGCTTCAGCGCTTCAGCGCTTCAGCGCTTCAGCGCTTCAG 560  
QY 54 ProGlyThrGly-----AlaSerAlaIleAlaAlaThrValThrProLysGlyAla 70  
Db 561 CTGGGCGCGGTTTCGTCTCGACCGCGCCCTGTGGTGGAGGCACTCGCGAGGCGG 620  
QY 71 SerMetLysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeu 90  
Db 621 CGGTCCGCGCTTCAGGTCACGAGACATGGGCGCCACACCCCGTGCCTCGACACCGC 680  
QY 91 SerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHis 110  
Db 681 GTTGGCGGT---CGCGAGGAGCAGCATGTGTGAGGTGCGCCCTGCACAGCGGACTGAA 737  
QY 111 ArgLeuPro-----GluGlyHisProLeuLeuGluLysArgAlaGluTyrPheArgHis 128

Db 738 CGAGTGCCTCTCGGTGAGACACCTCGGGTCTCGACGGCGCAGCATCCACGC--- 794  
 Qy 129 LeuArgSerLeuLysSerGlnGlyValAsnArg 139  
 Db 795 -----CTTCCACATCGAGGGCTCGGGCG 818  
 RESULT 32  
 US-09-920-923-1  
 ; Sequence 1, Application US/09920923  
 ; Publication No. US20030022273A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Pasamontes, Luis  
 ; APPLICANT: Tsygankov, Yuri  
 ; TITLE OF INVENTION: Improved Fermentative Carotenoid Production  
 ; FILE REFERENCE: Improved Fermentative Carotenoid  
 ; CURRENT APPLICATION NUMBER: US/09/920,923  
 ; CURRENT FILING DATE: 2001-08-02  
 ; PRIOR APPLICATION NUMBER: 08/980,832  
 ; PRIOR FILING DATE: 1997-12-01  
 ; NUMBER OF SEQ ID NOS: 66  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 8625  
 ; TYPE: DNA  
 ; ORGANISM: Flavobacterium sp. R1534  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (8348)..(8349)  
 ; NAME/KEY: unsure  
 ; LOCATION: (8539)..(8540)  
 ; NAME/KEY: unsure  
 ; LOCATION: (8581)  
 ; NAME/KEY: unsure  
 ; LOCATION: (8590)  
 ; NAME/KEY: unsure  
 ; LOCATION: (8592)  
 ; NAME/KEY: unsure  
 ; LOCATION: (8602)..(8604)  
 ; US-09-920-923-1  
 Alignment Scores:  
 Pred. No.: 64.1 Length: 8625  
 Score: 79.50 Matches: 35  
 Percent Similarity: 42.15% Conservative: 16  
 Best Local Similarity: 28.93% Mismatches: 50  
 Query Match: 11.23% Indels: 20  
 DB: 11 Gaps: 4  
 US-10-087-573-2 (1-141) x US-09-920-923-1 (1-8625)  
 Qy 37 ArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArg----- 51  
 Db 6353 CGCGTATCTCGATCAGGATCGGGTGGGACTGAAGGGCAGCAGATAGATGAAGCGGTAC 6412  
 Qy 52 ---ValLeuProGlyThrGlyAlaSerAlaIleAlaAlaThrValThrProLysGlyAla 70  
 Db 6413 CCGTCCATCTCGGAACGGTCGCTCCATGATCATCGGCGCTCGACGCCATGGGGCGG 6472  
 Qy 71 SerMetLysLeu-LysProProArg-----ProGlnSerThrLysSerProG1 86  
 Db 6473 TGGGTCGATCTCGACGCCACGAAATTTCTGGAACCCAGCGTCAGTCCGCGCGGTCTCG 6532  
 Qy 86 uLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAl 106  
 Db 6533 A---CGGCACACCGGCGTCGATCAGCAGGACGCTCGATCCGCGAGCGCTCGCTCAGC 6589  
 Qy 106 aArgValAsnHisArg-----LeuProGluGlyHisProLe 118  
 Db 6590 GTCCGCGCGGTATCGTCCAGTCGCGACATCGGTATTCACCGCAGATCGACACCTTCG 6649  
 Qy 118 uLeuGluLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAs 138  
 Db 6650 AGCAGCCCGATCAGCGCGCCCGCTCGATCGAGCCATAGCTGTCTGTCAGCGCGCGCAA 6709

Qy 138 n 138  
 Db 6710 T 6710  
 RESULT 33  
 US-09-920-923-27  
 ; Sequence 27, Application US/09920923  
 ; Publication No. US20030022273A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Pasamontes, Luis  
 ; APPLICANT: Tsygankov, Yuri  
 ; TITLE OF INVENTION: Improved Fermentative Carotenoid Production  
 ; FILE REFERENCE: Improved Fermentative Carotenoid  
 ; CURRENT APPLICATION NUMBER: US/09/920,923  
 ; CURRENT FILING DATE: 2001-08-02  
 ; PRIOR APPLICATION NUMBER: 08/980,832  
 ; PRIOR FILING DATE: 1997-12-01  
 ; NUMBER OF SEQ ID NOS: 66  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 27  
 ; LENGTH: 11233  
 ; TYPE: DNA  
 ; ORGANISM: Unknown  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Unknown Organism: Plasmid pZea4  
 ; US-09-920-923-27  
 Alignment Scores:  
 Pred. No.: 90.9 Length: 11233  
 Score: 79.50 Matches: 35  
 Percent Similarity: 42.15% Conservative: 16  
 Best Local Similarity: 28.93% Mismatches: 50  
 Query Match: 11.23% Indels: 20  
 DB: 11 Gaps: 4  
 US-10-087-573-2 (1-141) x US-09-920-923-27 (1-11233)  
 Qy 37 ArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArg----- 51  
 Db 7035 CGCGTATCTCGATCAGGATCGGGTGGGACTGAAGGGCAGCAGATAGATGAAGCGGTAC 7094  
 Qy 52 ---ValLeuProGlyThrGlyAlaSerAlaIleAlaAlaThrValThrProLysGlyAla 70  
 Db 7095 CCGTCCATCTCGGAACGGTCGCTCCATGATCATCGGCGCTCGACGCCATGGGGCGG 7154  
 Qy 71 SerMetLysLeu-LysProProArg-----ProGlnSerThrLysSerProG1 86  
 Db 7155 TCGGTCTCGATCTCGACGCCACGAAATTTCTGGAACCCAGCGTCAGTCCGCGGTCTCG 7214  
 Qy 86 uLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAl 106  
 Db 7215 A---CGGCACACCGGCGTCGATCAGCAGGACGCTCGATCCGCGAGCGCTCGCTCAGC 7271  
 Qy 106 aArgValAsnHisArg-----LeuProGluGlyHisProLe 118  
 Db 7272 GTCCGCGCGGTATCGTCCAGTCGCGACATGCGGTATTCACCGCAGATCGACACCTTCG 7331  
 Qy 118 uLeuGluLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAs 138  
 Db 7332 AGCAGCCCGATCAGCGCGCCCGCTCGATCGAGCCATAGCTGTCTGTCAGCGCGCGCAA 7391  
 Qy 138 n 138  
 Db 7392 T 7392  
 RESULT 34  
 US-10-156-761-6180  
 ; Sequence 6180, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO



```
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6180
; LENGTH: 756
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(756)
US-10-156-761-6180

Alignment Scores:
Pred. No.:      2.94      Length:      756
Score:          79.00     Matches:      32
Percent Similarity: 42.05% Conservative:    5
Best Local Similarity: 36.36% Mismatches:     37
Query Match:      11.16% Indels:       14
DB:               14      Gaps:         4

US-10-087-573-2 (1-141) x US-10-156-761-6180 (1-756)
QY      6 ThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMet 25
Ddb      461 TCTACTACAGTCGCGGAGCGGCAGCGCCGAACTGCCGGCGCACGGTCTGGTCTGGA 520
QY      26 ArgGluAlaLeuLeuArgValLySerSerGluArgLeuAlaMetLeuArgAla----- 43
Ddb      521 AGGTGCTACC GGCGCGCG-- --AATCCACGAGCGCGCTCGGAGCTGCCCGCAGATCG 577
QY      44 -----LeuAlaGlyMet-CysGlyHisArgValLeuProGlyThrGlyAla----- 58
Ddb      578 CGCCAAGACGCGCGCGTCTGCGGTTCGGGAAGCGCCATCAACGGGATCGACCCGG 637
QY      59 ---SerAlaIleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProPro 77
Ddb      638 TGGACGTCGCGCGCAGCTACC----- -GCTTCGACGAGCGGCTTCACCTTCG 682
QY      78 ArgProGlnSerThrLysSerPro 85
Ddb      683 AGGCCGGCCTCGACGGGTGCGCG 706

RESULT 35
US-10-156-761-6333
; Sequence 6333, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
```

```
; SEQ ID NO 1558
; LENGTH: 1860
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1860)
US-10-156-761-1558

Alignment Scores:
Pred. No.: 9.69 Length: 1860
Score: 79.00 Matches: 39
Percent Similarity: 40.31% Conservative: 13
Best Local Similarity: 30.23% Mismatches: 41
Query Match: 11.16% Indels: 36
DB: 14 Gaps: 8

US-10-087-573-2 (1-141) x US-10-156-761-1558 (1-1860)
QY 4 ThrSerThrThrThrAsnPheValAla-----GluAsn 14
Db 488 ACACAGACCACACGACGGTTCACGGCGCGCGGGTGAACCGGAGACGGCCCTCGACGAAC 429
QY 15 ArgProThrPheCly-----GluThrPheAspValMetArgGluAlaLeu 30
Db 428 CAGCCGATCTTCGACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 369
QY 31 ArgValLysSerSerGluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHis 50
Db 368 CGGGTCAGGCGGAGCGGAGCTCGCGGGTCTCGCG-----GCCAGCAGCTGC----- 321
QY 51 ArgValLeuProGlyThrGlyAlaSerAlaAlaAlaThrValThrProLysGlyAla 70
Db 320 -----TGGACGCGCGCGCGCGCGCGGTATCCAGTTGGTGACGCGCGCGCGCGGTGC 273
QY 71 SerMetLysLeuLysPro-----ProArgProGlnSerThrLysSerProGluLeu 87
Db 272 AGCATGGAT--TCGCCCCAGGCGGAGCGCGCGCGCGCGCGCACAGCAGCATCAGCCCGCGGTC 216
QY 88 ArgGluLeuSerArgLysIleArgGluMetAsnLysThrIle----- 101
Db 215 AGG-----GCGAGTCGGGTGCCGAGCGGTCCACCACCGCTGGACGCGGCGGCGGACG 162
QY 102 -----SerGlnGluSerAlaArg 107
Db 161 CAGACCGCAGCCCGGAAACGGCCAGG 135

RESULT 37
US-09-864-636A-431
; Sequence 431, Application US/09864636A
; Publication No. US20030104378A1
; GENERAL INFORMATION:
; APPLICANT: Third Wave Technologies
; APPLICANT: Allwai, Hatim
; APPLICANT: Bartholomay, Christian
; APPLICANT: Chehak, LuAnne
; APPLICANT: Eis, Peggy S.
; APPLICANT: Hall, Jeff G.
; APPLICANT: Ip, Hon S.
; APPLICANT: Ji, Lin
; APPLICANT: Kaiser, Michael
; APPLICANT: Kwiatkowski, Jr., Robert W.
; APPLICANT: Lukowski, Andrew A.
; APPLICANT: Lyamichev, Victor
; APPLICANT: Lymaicheva, Natalie E.
; APPLICANT: Ma, WuPo
; APPLICANT: Neri, Bruce P.
; APPLICANT: Olson, Sarah M.
; APPLICANT: Olson-Munoz, Marilyn C.
; APPLICANT: Schaefer, James J.
; APPLICANT: Skrzypczynski, Zbigniew
; APPLICANT: Takova, Teetska Y.
; APPLICANT: Thompson, Lisa C.
; APPLICANT: Vedvik, Kevin L.
; TITLE OF INVENTION: RNA Detection Assays

US-09-864-636A-431
Alignment Scores:
Pred. No.: 14.4 Length: 2508
Score: 79.00 Matches: 54
```

FILE REFERENCE: FORS-06666  
CURRENT APPLICATION NUMBER: US/10/084,839  
CURRENT FILING DATE: 2002-02-26  
NUMBER OF SEQ ID NOS: 4004  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 431  
LENGTH: 2508  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Synthetic  
US-10-084-839-431  
Alignment Scores:  
Pred. No.: 14.4 Length: 2508  
Score: 79.00 Matches: 54  
Percent Similarity: 42.13% Conservatives: 21  
Best Local Similarity: 30.34% Mismatches: 58  
Query Match: 11.16% Indels: 46  
Gaps: 10  
US-10-087-573-2 (1-141) x US-10-084-839-431 (1-2508)  
Qy 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20  
Db 1150 CTGGACCCCTCGAACACACCCCGAGGGGTGGCG-----CGCGCTACGGGGGGAG 1203  
Qy 21 ---ThrPheAspValMetArgGluAlaLeu----- 30  
Db 1204 TGGACGAGGACCGCCCGACCGGCCCTCTCTCGGAGAGGCTCCATCGGAACCTCCTT 1263  
Qy 31 ---ArgValLysSerSerGluArgLeuAlaMetLeu----- 41  
Db 1264 AAGCGCTTCGAGGGGAGGAGAGCTCTCTGGCTCTACACGAGGTGGAAAAGCCCTC 1323  
Qy 42 ---ArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGly----- 57  
Db 1324 TCCCGGCTCTGGCCCATATGGA-GGCCACCGGGGTACGGGGGACGTGGCTTACCTTCA 1382  
Qy 58 -----AlaSerAlaIleAlaAlaThrValThrProLysGlyAlaSerMetLysLeu 74  
Db 1383 GCGCCCTTCCCTGGAGCTTGGCGGAGAGATCGCGGCTCGAGGAGGAGGTCTTCGGCTT 1442  
Qy 75 LysProArgProGlnSerThrLysSerProGluLeu-Arg---GluLeuSerArg-- 92  
Db 1443 GCGGGGCCACCCCTTCAA-----CTCAACTCCCGGGACCCAGCTGGAAAGGTT 1490  
Qy 93 -----LysIleArgGluMetAsnLysThrIleSerGlnG 104  
Db 1491 GCTCTTGACGAGCTTAGGCTTCCCAAGTTGAAGAAGACGAAGAAGACAGGCAAGCGCTC 1550  
Qy 104 u---SerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAl 123  
Db 1551 CACCAGCGCGGGTGTGGAGGCCCTACGGAGGCCACCCCATCGTGGAGAAGATCCT 1610  
Qy 123 aGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140  
Db 1611 CCAGCAC---CGGAGCTCACCAGCTCAAGACACCTACGTGGAGCCCTC 1659  
RESULT 39  
US-09-864-636A-73  
Sequence 73, Application US/09864636A  
Publication No. US20030104378A1  
GENERAL INFORMATION:  
APPLICANT: Third Wave Technologies  
APPLICANT: Allwai, Hatim  
APPLICANT: Bartholomay, Christian  
APPLICANT: Chehak, LuAnne  
TITLE OF INVENTION: Detection of RNA Sequences  
FILE REFERENCE: FORS-04944  
CURRENT APPLICATION NUMBER: US/09/864,636A  
CURRENT FILING DATE: 2002-10-15  
NUMBER OF SEQ ID NOS: 2640

SOFTWARE: PatentIn version 3.0  
SEQ ID NO 73  
LENGTH: 2517  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Synthetic  
US-09-864-636A-73  
Alignment Scores:  
Pred. No.: 14.5 Length: 2517  
Score: 79.00 Matches: 54  
Percent Similarity: 42.13% Conservatives: 21  
Best Local Similarity: 30.34% Mismatches: 58  
Query Match: 11.16% Indels: 46  
Gaps: 10  
US-10-087-573-2 (1-141) x US-09-864-636A-73 (1-2517)  
Qy 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20  
Db 1141 CTGGACCCCTCGAACACACCCCGAGGGGTGGCG-----CGGCGCTACGGGGGGAG 1194  
Qy 21 ---ThrPheAspValMetArgGluAlaLeu----- 30  
Db 1195 TGGACGAGGACCGCCCGACCGGCCCTCTCTCGGAGAGGCTCCATCGGAACCTCCTT 1254  
Qy 31 ---ArgValLysSerSerGluArgLeuAlaMetLeu----- 41  
Db 1255 AAGCGCTTCGAGGGGAGGAGAGCTCTCTGGCTCTACACGAGGTGGAAAAGCCCTC 1314  
Qy 42 ---ArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla 60  
Db 1315 TCCCGGCTCTGGCCCATATGGA-GGCCACCGGGGTGGCGCTCGAGCTGGC----- 1364  
Qy 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProArg----- 78  
Db 1365 CTATCTCAGGCGCTTGTCCCT---GGAGGTGGCGGAGGAGATCGCCGCTCGAGGCCGA 1421  
Qy 79 -----ProGln-----SerThrLysSer 84  
Db 1422 GGTCTTCGCGCTCGCGGCCACCCCTTCAACCTCCCGGAGCACAGCTGGAAAGGTT 1481  
Qy 85 ProGluLeuArgGluLeuSerArgLysIleArgGlu-MetAsnLysThrIleSerGlnG 104  
Db 1482 CTTCTTGACGAGCTAGGGCTTCCGCCATCGCAAGACGAGAGACCGGCAAGCGCTC 1541  
Qy 104 u---SerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAl 123  
Db 1542 CACCAGCGCGCGTCTCGGAGGCCCTCGCGAGGCCACCCCATCGTGGAGAAGATCCT 1601  
Qy 123 aGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140  
Db 1602 CGAGTAC---CGGAGCTCACCAGCTGAAGAGACCTACATTGACCCCTTG 1650  
RESULT 40  
US-09-864-636A-87  
Sequence 87, Application US/09864636A  
Publication No. US20030104378A1  
GENERAL INFORMATION:  
APPLICANT: Third Wave Technologies  
APPLICANT: Allwai, Hatim  
APPLICANT: Bartholomay, Christian  
APPLICANT: Chehak, LuAnne  
TITLE OF INVENTION: Detection of RNA Sequences  
FILE REFERENCE: FORS-04944  
CURRENT APPLICATION NUMBER: US/09/864,636A  
CURRENT FILING DATE: 2002-10-15  
NUMBER OF SEQ ID NOS: 2640  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 87  
LENGTH: 2517  
TYPE: DNA

```

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-864-636A-87

```

```

Alignment Scores:
Pred. No.: 14.5 Length: 2517
Score: 79.00 Matches: 54
Percent Similarity: 42.13% Conservative: 21
Best Local Similarity: 30.34% Mismatches: 58
Query Match: 11.16% Indels: 46
DB: 11 Gaps: 10

```

US-10-087-573-2 (1-141) x US-09-864-636A-87 (1-2517)

```

QY 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20
DB 1141 CTGGACCTTCGAACACCCAGGGGGTGGCG-----CGGCGTACGGGGGGAG 1194
QY 21 ---ThrPheAspValMetArgGluAlaLeu----- 30
DB 1195 TGGACGGAGGACGCCGCCACCGGGCCCTCTCTCGAGAGGCTCCATCGGAAGCTCTCT 1254
QY 31 ---ArgValLysSerSerGluArgLeuAlaMetLeu----- 41
DB 1255 AAGCCCTCGAGGGGAGGAGAGCTCTTTGGCTCTACACGAGGTGGAAAGCCCTC 1314
QY 42 ---ArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla 60
DB 1315 TCCCGGCTCTGCCCATATGA-GGCCACGGGGGTGGCGGACGTGGC----- 1364
QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArg----- 78
DB 1365 CTATCTCAGGGCTTGTCCCT---GGAGGTGGCCGAGGAGATCGCCCGCTCGAGGCCGA 1421
QY 79 -----ProGln-----SerThrLysSer 84
DB 1422 GGTCTTCGCGCTGGCGGCCACCCCTTCAACCTCACTCCCGGACACAGCTGGAAGGT 1481
QY 85 ProGluLeuArgGluLeuSerArgLysIleArgGlu-MetAsnLysThrIleSerGlnGln 104
DB 1482 CCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACGGAGAACCGGCAAGCGCTC 1541
QY 104 u---SerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAl 123
DB 1542 CACCACGGCGCGCTCTCGAGGCCCTCGCGAGGCCACCCCTCATCGTGAGAGATCCT 1601
QY 123 aGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140
DB 1602 GCAGTAC---CGGGAGCTCACCAGCTGAAGAGACCTACATTGACCCCTTG 1650

```

Search completed: November 17, 2003, 15:54:59  
Job time : 3118 secs

29: gb\_gss2:\*

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 17, 2003, 12:52:44 ; Search time 1928 Seconds  
(without alignments)  
1777.454 Million cell updates/sec

Title: US-10-087-573-2  
Perfect score: 708  
Sequence: 1 MESTSTTTNFVAENRPTFG.....RAEYFRLHSLKSGQVRLI 141

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Zgapop 6.0 , Zgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model DEV=xlh  
-Q=/cgn2/1/USPRO.spool/US10087573/runat.14112003.103610.20685/app\_query.fasta.1.327  
-DB=EST -QMT=fastap -SUFFIX=ret -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-OCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10087573 @CGN 1 1 2135 @runat.14112003.103610.20685 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIME=120 -WARN\_TIME=30 -THREAS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hcc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_est3:\*  
12: gb\_est4:\*  
13: gb\_est5:\*  
14: gb\_est6:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	101.5	14.3	665	14	CD468520	LeukoS3_3
C 2	96.5	13.6	1469	29	AG065308	Pan trogl
C 3	95.5	13.5	431	12	BI961232	MONOL 7_G
C 4	95.5	13.5	574	14	CD467780	LeukoS1_6
C 5	95.5	13.5	646	14	CD536579	LeukoN8
C 6	95.5	13.5	654	10	BE513677	601315412
C 7	95.5	13.5	663	14	CD468466	LeukoS3_3
C 8	95.5	13.5	671	14	CD467297	LeukoS1_3
C 9	95.5	13.5	700	14	CD468492	LeukoS3_3
C 10	95.5	13.5	729	14	CD535377	LeukoN5_2
C 11	95.5	13.5	738	14	CD528498	LeukoN3_2
C 12	95	13.4	881	12	BI248636	LeukoS3_3
C 13	94.5	13.3	496	14	CD470400	LeukoS4_4
C 14	94	13.3	607	10	BI180114	602329694
C 15	94	13.3	669	12	BI181817	603033383
C 16	94	13.3	927	29	CC366660	PUHSH80TD
C 17	94	13.3	1172	12	BM909493	AGENCOURT
C 18	93.5	13.2	660	14	BY749167	BY749167
C 19	92.5	13.1	554	9	AV667883	AV667883
C 20	92.5	13.1	578	9	AV597776	AV597776
C 21	92.5	13.1	578	9	AV616040	AV616040
C 22	92.5	13.1	594	9	AV601755	AV601755
C 23	92.5	13.1	615	10	BE371432	RC0-FN014
C 24	92.5	13.1	641	10	BE574311	SNBST4312
C 25	92.5	13.1	667	12	BP112124	BP112124
C 26	92.5	13.1	691	12	BI771910	603058924
C 27	92.5	13.1	875	12	BI757298	603029512
C 28	92.5	13.1	907	10	BF301271	602029713
C 29	92	13.0	1104	29	CNS03BBW	AL236309 Tetraodon
C 30	91.5	12.9	425	13	BY224833	BY224833
C 31	91.5	12.9	544	10	BE553346	BE553346
C 32	91.5	12.9	640	10	BE284887	BE284887
C 33	91.5	12.9	650	10	BE289370	BE289370
C 34	91.5	12.9	674	12	BI150185	602848591
C 35	91.5	12.9	727	10	BF581991	BF581991
C 36	91.5	12.9	768	12	BI696433	60345762
C 37	91.5	12.9	786	12	BI158106	60320741
C 38	91.5	12.9	835	12	BI833165	603086487
C 39	91.5	12.9	849	12	BI411683	BI411683
C 40	91.5	12.9	872	14	CB586950	CB586950
C 41	91.5	12.9	900	13	BQ952740	BQ952740
C 42	91.5	12.9	929	10	BF302274	BF302274
C 43	91.5	12.9	934	12	BI525023	602926430
C 44	91.5	12.9	961	12	BI904618	BI904618
C 45	91.5	12.9	968	12	BI108132	BI108132

ALIGNMENTS

RESULT 1  
LOCUS CD468520/c  
DEFINITION LeukoS3\_3\_D04.g1\_A025 Stimulated peripheral blood leukocytes S3  
ACCESSION CD468520  
VERSION CD468520  
KEYWORDS EST.  
SOURCE Equus caballus (horse)  
ORGANISM Equus caballus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
REFERENCE 1 (bases 1 to 665)

AUTHORS Vandenplas,M., Cordonnier-Pratt,M.-M., Suzuki,Y., Sugano,S., Moore  
J.N., Liang,C., Sun,F., Sullivan,R., Shah,M. and Pratt,L.H.  
TITLE An EST database from equine (Equus caballus) stimulated peripheral  
blood leukocytes  
JOURNAL Unpublished  
COMMENT Other\_ESTS: LeukoS3\_3\_D04\_b1\_A025  
Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu  
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in  
the Human Genome Center, University of Tokyo Institute of Medical  
Science; tissue and RNA were prepared in the department of Large  
Animal Medicine, University of Georgia; sequencing done in the  
Laboratory for Genomics and Bioinformatics, University of Georgia.  
Sequence ends have been trimmed to exclude vector and regions below  
phred quality 16. Three-prime sequences are presented as their  
reverse complement and have been trimmed to exclude polyA.  
Seq primer: Sug5 (CTCTGCTCTAAAGCTGGC).

FEATURES  
Location/Qualifiers

1..665  
/organism="Equus caballus"  
/mol\_type="mRNA"  
/strain="Tennessee walking horse"  
/db\_xref="taxon:9796"  
/clone="LeukoS3\_3\_D04\_A025"  
/sex="female"  
/tissue\_type="blood"  
/cell\_type="leukocytes"  
/lab\_host="DH10B-T1 phage-resistant E. coli"  
/note="Organ: Stimulated peripheral blood leukocytes S3"  
/clone\_lib="Stimulated peripheral blood leukocytes S3"  
Site 1: XhoI; Site 2: XhoI; The library was prepared from  
polyA+ RNA from equine peripheral blood leukocytes  
isolated from a healthy adult horse. The leukocytes were  
stimulated for 4 hr with 10 ng/ml E. coli O55:B5 LPS.  
Double-stranded cDNA was cloned unidirectionally into  
different DraIII sites of the pME18S-FL3 vector (5-prime  
DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG  
). XhoI excises the cDNA insert."

BASE COUNT 116 a 200 c 210 g 139 t

ORIGIN

Alignment Scores:  
Pred. No.: 3.79 Length: 665  
Score: 101.50 Matches: 40  
Percent Similarity: 44.4% Conservative: 16  
Best Local Similarity: 31.7% Mismatches: 46  
Query Match: 14.3% Indels: 24  
DB: 14 Gaps: 5

US-10-087-573-2 (1-141) x CD468520 (1-665)

QY 3 SerThrSerThrThrThrThrThrValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
Db 491 TCCAGGGTGACGACAGCTTCCTCAGGGGCGACACCGCCCGCTCAGGATCTCCAC 432  
QY 23 AspValMetArgGluAlaLeuLeuArgVallyssSerGluArgLeuAlaMetLeuArg 42  
Db 431 -----ACCATGACTGTGCAATCTCTCGAGCCCACTGGCAATGACGTTA 390  
QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
Db 389 TCATTG-----TGGGGCAGCAGCGATGTCCACACAGGGGCTGTGTGCCACAG 339  
QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
Db 338 ACCGTGGGGCGATCTTGTCCACACGTCACAGTCTTGTCCCGGGGCGAGCACCAGGAAGGCC 279  
QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94

Db 278 CTCCCGCGCTGGCTCCATACGAGGGCCGAACTTGGGGTTGACGACACAG----- 225  
QY 95 ArgGluMetAsnLysThrIleSerGlnGlu-SerAlaArgValAsnHisArgLeuProGln 114  
Db 224 -----AAGCCACTGTCCAGGTGGTCTGGAGACGCGGCACATCTCTCATACAC 177  
QY 114 uGlyHisProLeuLeu 119  
Db 176 TGGTGGCGCTTGGCTG 161

RESULT 2  
AG065308

LOCUS AG065308 1469 bp DNA linear GSS 03-NOV-2001  
DEFINITION Pan troglodytes DNA, clone: PTB-054K10.R, genomic survey sequence.  
ACCESSION AG065308  
VERSION AG065308.1 GI:16617110  
KEYWORDS GSS.  
SOURCE Pan troglodytes (chimpanzee)  
ORGANISM Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

REFERENCE 1

AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE BAC end sequences of Library PTB

JOURNAL

REFERENCE 2 (bases 1 to 1469)

AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE Direct Submission

JOURNAL

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail:chimbases@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/  
Tel:81-45-503-9111, Fax:81-45-503-9170)

Clones are derived from the chimpanzee BAC library PTB This BAC end  
was generated during the R&D process and may have higher chance of  
clone tracking errors.

PRIMERS

Sequencing: M13Rev

LIBRARY

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI

FEATURES

source

1..1469  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/clone="PTB-054K10.R"  
/sex="male"  
/cell\_type="lymphoblast"  
/clone\_lib="PTB Chimpanzee Male BAC Library"

BASE COUNT 313 a 591 c 311 g 233 t

ORIGIN

Alignment Scores:

Pred. No.: 32.6 Length: 1469

Score: 96.50 Matches: 36

Percent Similarity: 40.44% Conservative: 19

Best Local Similarity: 26.47% Mismatches: 46

Query Match: 13.63% Indels: 35

DB: 29 Gaps: 6

US-10-087-573-2 (1-141) x AG065308 (1-1469)

QY 12 AlaGluAsnArgProThr-----PheGlyGluThrPheAspVal 24  
Db 713 GCTCAGACCGCCCGAGCTACCTCAATCGCTACGACCCAGCACACTGC 772  
QY 25 MetArgGluAlaLeuArgVallyssSerGluArgLeuAlaMetLeuArgAlaLeu 44  
Db 25 MetArgGluAlaLeuArgVallyssSerGluArgLeuAlaMetLeuArgAlaLeu 44

Db 773 CCACGCTACCACTACTCAAAAGTATGATGACGAGCAGCTTCTGCATGACTCCGAAGTGTG 832

Qy 45 AlaGlyMetCysGlyHisArgValLeuProGly 56

Db 833 GCG-----CGTCTCACTCCCTCGATGAGTGCCTTCGATCTCCACGCC 877

Qy 57 GlyAlaSerAlaIleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysPro 76

Db 878 GCGCGCGTGTCTCGCGGCTGCATCCGCCGCTACGTCACCTGCGC---AGCACC 934

Qy 77 ProArgProGlnSerThrLysSerProGluLeu-----ArgGluLeuSerArgLys 93

Db 935 CCGCGCCCTCTCTCCGCCAGTCGTCCTCGACTCCACCGATCTCGTCCGCTTCACGACGN 994

Qy 94 IleArgGluMet-----AsnLysThrIleSer 102

Db 995 GTACACCATCTTACCATCACATCCGTCGCTCGATACGACCTCCACACACGCGCATCC 1054

Qy 103 GlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeu 118

Db 1055 TCGCAATCTTACGCGCTAACCTGCGCTACCGCGCGCGGCGGCGGACCACTA 1102

RESULT 3

LOCUS BI961232/c 431 bp mRNA linear EST 22-OCT-2001

DEFINITION MONO1.7\_G02.bi\_A005 Monocytes (MONO1) Equus caballus cDNA, mRNA

ACCESSION BI961232.1 GI:16319435

VERSION BI961232

KEYWORDS EST.

SOURCE Equus caballus (horse)

ORGANISM Equus caballus

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Perissodactyla; Equidae; Equus.

AUTHORS Vandenplas,M., Cordonnier-Pratt,M.-M., Sudman,M.L., Wentzel,V.E., Gingle,A.R., Pratt,L.H. and Moore,J.N.

TITLE An EST database from equine (Equus caballus) monocytes

JOURNAL Unpublished

COMMENT Contact: Cordonnier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 583 0210 Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for high quality sequence is 20. Three-prime sequences, which are obtained with PolyMix or T7 sequencing primer, are presented as the reverse complement.

Seq primer: JEN REV

High quality sequence stop: 425

POLYA=No.

FEATURES

source

1..431 Location/Qualifiers

/organism="Equus caballus"

/mol\_type="mRNA"

/db\_xref="taxon:9796"

/cell\_type="isolated peripheral blood monocytes stimulated with E. coli lipopolysaccharide"

/clone\_lib="Monocytes (MONO1)"

/note="Vector: pBluescript SK(-) from Lambda ZapII; Site 1: XhoI; Site 2: EcoRI; The library was made from poly(A) RNA in the cloning vector lambda ZapII. Clones to be sequenced were prepared by mass excision."

BASE COUNT 78 a 127 c 147 g 79 t

ORIGIN

Alignment Scores:

Pred. No.: 8.52 Length: 431

Score: 95.50 Matches: 39

Percent Similarity: 43.65% Conservative: 16

Best Local Similarity: 30.95% Mismatches: 47

Query Match: 13.49% Indels: 24

DB: 12 Gaps: 5

US-10-087-573-2 (1-141) x BI961232 (1-431)

Qy 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22

Db 349 TCCAGGTTGACGACAGCTTCCCTCAGGCGGAGCAGCAGGCCCCCGTCAGGGATCTCCAC 290

Qy 23 AspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42

Db 289 -----ACCATGACTGTGCAATCTCCGAGCCACTGGCAATGACGTTA 248

Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58

Db 247 TCATTG-----TCGCGGCACAGGCGCATGTCAGCAGCAGGGCTGTGTGGCCACAG 197

Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74

Db 196 ACGTGGCGCATCTTGTCCACAGTCCTGTCGCCAGGCGGAGCAGCAGGAGGCC 137

Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94

Db 136 CTTCCCGCGTGGCCTCACATATGAGGCGCCACGAATCTGGGGTTGACAGCAG----- 83

Qy 95 ArgGluMetAsnLysThrIleSerGlnGlu-SerAlaArgValAsnHisArgLeuProGln 114

Db 82 -----AAGCACGTCTCCAGTGGTCTGCGAGAGCGGCACATCTCTCATAACAC 35

Qy 114 uGlyHisProLeuLeu 119

Db 34 TGGTCGCGCTTGGCTG 19

RESULT 4

LOCUS CD467780/c 574 bp mRNA linear EST 04-JUN-2003

DEFINITION LeukoS1\_6\_E02.g1\_A023 Stimulated peripheral blood leukocytes S1

ACCESSION CD467780

VERSION CD467780.1 GI:31389048

KEYWORDS EST.

SOURCE Equus caballus (horse)

ORGANISM Equus caballus

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Perissodactyla; Equidae; Equus.

AUTHORS Vandenplas,M., Cordonnier-Pratt,M.-M., Suzuki,Y., Sugano,S., Moore,J.N., Liang,C., Sun,F., Sullivan,R., Shah,M. and Pratt,L.H.

TITLE An EST database from equine (Equus caballus) stimulated peripheral blood leukocytes

JOURNAL Unpublished

COMMENT Other ESTs: LeukoS1\_6\_E02.b1\_A023 Contact: Cordonnier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 583 0210 Email: mmpratt@uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; tissue and RNA were prepared in the Department of Large Animal Medicine, University of Georgia; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.

Seq primer: Sug5 (CTTCTGCTCTAAAGCTCG).

FEATURES

source

1..574 Location/Qualifiers

/organism="Equus caballus"

/mol\_type="mRNA"

/strain="Belgian draft"

/db\_xref="taxon:9796"  
/clone="LeukOS1\_6\_E02\_A023"  
/sex="female"  
/tissue\_type="blood"  
/cell\_type="leukocytes"  
/lab\_host="DH10B-T1 phage-resistant E. coli"  
/clone\_lib="Stimulated peripheral blood leukocytes S1"  
/note="Organ: circulatory system; Vector: pME18S-FL3;  
Site 1: XhoI; Site 2: XhoI; The library was prepared from  
polyA+ RNA from equine peripheral blood leukocytes  
isolated from a healthy adult horse. The leukocytes were  
stimulated for 4 hr with 10 ng/ml E. coli O55:B5 LPS.  
Double-stranded cDNA was cloned unidirectionally into  
different DraIII sites of the pME18S-FL3 vector (5-prime  
DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG  
). XhoI excises the cDNA insert."  
BASE COUNT 99 a 179 c 176 g 120 t  
ORIGIN

Alignment Scores:  
Pred. No.: 12.3 Length: 574  
Score: 95.50 Matches: 39  
Percent Similarity: 43.65% Conservative: 16  
Best Local Similarity: 30.95% Mismatches: 47  
Query Match: 13.49% Indels: 24  
DB: 14 Gaps: 5

US-10-087-573-2 (1-141) x CD467780 (1-574)

QY 3 SerThrSerThrThrAsnPhenValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
Db 486 TCCAGGGTGACGACAGCTTCCCTCAGGGGACGACCCAGGGCCCGCTCAGGGATCTCCAC 427  
QY 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
Db 426 -----ACCATGACTGTGCAATCTCCGAGCCACTGCGCAATGACGTTA 385  
QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
Db 384 TCATTG-----TCCGGGACACGAGGATCTCCAGCAGCAGGGGCTGTGGGCGACAG 334  
QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
Db 333 ACCGTGGGCGCATTTGTCACACGTCAGTCTTCCAGGGGACGACCGAGGAGGCC 274  
QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
Db 273 CTTCCCGCGCTGGCCTCACATATGAGGGCCAGCACTTGGGGTTGACAGCAG----- 220  
QY 95 ArgGluMetAsnLysThrIleSerGlnGlu-SerAlaArgValAsnHisArgLeuProG 114  
Db 219 -----AAGCCACTGTCAGGTGCTGCGAGCGGCACATCTCTCATTAACAC 172  
QY 114 uGlyHisProLeuLeu 119  
Db 171 TGTGCGGCTTGGCTG 156

RESULT 5  
CD536579/c 646 bp mRNA linear EST 10-JUN-2003  
LOCUS  
DEFINITION  
LeukOS1\_6\_E02\_A028 Unstimulated peripheral blood leukocytes N6  
EQUUS caballus cDNA clone LeukOS1\_6\_E02\_A028 5', mRNA sequence.  
ACCESSION  
CD536579  
VERSION  
EST.  
KEYWORDS  
CD536579.1 GI:31578994  
SOURCE  
EQUUS caballus (horse)  
ORGANISM  
EQUUS caballus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Perissodactyla; Equidae; EQUUS.  
REFERENCE  
1 (bases 1 to 646)  
AUTHORS  
Vandenplas, M., Cordonnier-Pratt, M.-M., Suzuki, Y., Sugano, S., Moore  
J.N., Liang, C., Sun, F., Sullivan, R., Shah, M., and Pratt, L.H.  
An EST database from equine (EQUUS caballus) unstimulated

peripheral blood leukocytes  
Unpublished  
Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu  
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in  
The Human Genome Center, University of Tokyo Institute of Medical  
Science; tissue and RNA were prepared in the Department of Large  
Animal Medicine, University of Georgia; sequencing done in the  
Laboratory for Genomics and Bioinformatics, University of Georgia.  
Sequence ends have been trimmed to exclude vector and regions below  
phred quality 16. Three-prime sequences are presented as their  
reverse complement and have been trimmed to exclude polyA.  
Seq primer: Sug5 (CTTGTGCTTAAAGACTGCG).  
Location/Qualifiers

FEATURES  
source

1..646  
/organism="EQUUS caballus"  
/mol\_type="mRNA"  
/strain="quarternhorse"  
/db\_xref="taxon:9796"  
/clone="LeukOS1\_6\_E02\_A028"  
/sex="male"  
/tissue\_type="blood"  
/cell\_type="leukocytes"  
/lab\_host="DH10B-T1 phage-resistant E. coli"  
/clone\_lib="Unstimulated peripheral blood leukocytes N6"  
/note="Organ: circulatory system; Vector: pME18S-FL3;  
Site 1: XhoI; Site 2: XhoI; The library was prepared from  
polyA+ RNA from unstimulated equine peripheral blood  
leukocytes isolated from a healthy adult horse.  
Double-stranded cDNA was cloned unidirectionally into  
different DraIII sites of the pME18S-FL3 vector (5-prime  
DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG  
). XhoI excises the cDNA insert."  
BASE COUNT 113 a 195 c 202 g 136 t  
ORIGIN

Alignment Scores:  
Pred. No.: 14.3 Length: 646  
Score: 95.50 Matches: 39  
Percent Similarity: 43.65% Conservative: 16  
Best Local Similarity: 30.95% Mismatches: 47  
Query Match: 13.49% Indels: 24  
DB: 14 Gaps: 5

US-10-087-573-2 (1-141) x CD536579 (1-646)

QY 3 SerThrSerThrThrAsnPhenValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
Db 491 TCCAGGGTGACGACAGCTTCCCTCAGGGGACGACCCAGGGCCCGCTCAGGGATCTCCAC 432  
QY 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
Db 431 -----ACCATGACTGTGCAATCTCCGAGCCACTGCGCAATGACGTTA 390  
QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
Db 389 TCATTG-----TCCGGGACACGAGGATCTCCAGCAGCAGGGGCTGTGTGGCCACAG 339  
QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
Db 338 ACCGTGGGCGCATTTGTCACACGTCAGTCTTCCAGGGGACGACCGAGGAGGCC 279  
QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
Db 278 CTTCCCGCGCTGGCCTCACATATGAGGGCCAGCACTTGGGGTTGACAGCAG----- 225  
QY 95 ArgGluMetAsnLysThrIleSerGlnGlu-SerAlaArgValAsnHisArgLeuProG 114



[illegible]

QY	59	-----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu	74
DB	305	ACCGTGGGGCGGCAATCTTGTCACACAGTCCAGTCTTGCCCGGGGACAGCAGGAGGCC	246
QY	75	LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle	94
DB	245	CCTGCCCGCG-----CTGGCCTCACAGATC	222
QY	95	ArgGluMetAsn-----LysThrIleSerGlnGlu-SerAlaArgVa	108
DB	221	AGGGCCACAACTTAGGTTGACAGCACAGAAGCCACTGTCAGGTGGTCTGTGTGAGACG	162
QY	108	IaslnHisArgLeuProGluGlyHisProLeu-----Le	119
DB	161	CCCAATCTTCATAGCACTGGTCGGCCATGGCCGGCTCTCAACACGTCGGCGGAACCT	102
QY	119	uGluLysArgAlaGluTyPheArgHisLeuArgSerLeuLysSerGln	135
DB	101	GCTGGAGCGGACCACTCGCGCGCTATTCTGTCGCCGAGAGGACAG	53
RESULT	7		
LOCUS	CD468466/c		
DEFINITION	LeukOS3_3_G11_g1_A025 Stimulated peripheral blood leukocytes S3		
ACCESSION	CD468466		
VERSION	1		
KEYWORDS	CD468466.1 GI:31389734		
SOURCE	Equus caballus (horse)		
ORGANISM	Equus caballus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Perissodactyla; Equidae; Equus.		
AUTHORS	1 (bases 1 to 663) Vandenplas, M., Cordonnier-Pratt, M.-M., Suzuki, Y., Sugano, S., Moore, J.N., Liang, C., Sun, F., Sullivan, R., Shah, M. and Pratt, L.H.		
TITLE	An EST database from equine (Equus caballus) stimulated peripheral blood leukocytes		
JOURNAL	Unpublished		
COMMENT	Other ESTs: LeukOS3_3_G11.bi.A025 Contact: Cordonnier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 583 0210 Email: mmpratt@uga.edu Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science, tissue and RNA were prepared in the Department of Large Animal Medicine, University of Georgia; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA. Seq primer: Sug5 (CTTCTGCTCTAAAAGCTCG).		
FEATURES	Location/Qualifiers		
source	1. .563 /organism="Equus caballus" /mol_type="mRNA" /strain="Tennessee walking horse" /db_xref="taxon:9796" /clone="LeukOS3_3_G11_A025" /sex="female" /tissue_type="blood" /cell_type="leukocytes" /lab_host="DH10B-T1 phage-resistant E. coli" /clone_lib="Stimulated peripheral blood leukocytes S3" /note="Organ: circulatory system; Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from polyA+ RNA from equine peripheral blood leukocytes isolated from a healthy adult horse. The leukocytes were stimulated for 4 hr with 10 ng/ml P. coli 055:85 LPS		

Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5'-prime DraIII site is CACTGTGTG, 3'-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."

BASE COUNT 118 a 198 c 205 g 142 t

ORIGIN

Alignment Scores:  
Pred. No.: 14.8 Length: 663  
Score: 95.50 Matches: 39  
Percent Similarity: 43.65% Conservative: 16  
Best Local Similarity: 30.95% Mismatches: 47  
Query Match: 13.49% Indels: 24  
DB: 14 Gaps: 5

US-10-087-573-2 (1-141) x CD468466 (1-663)

Qy 3 SerThrSerThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
Db 501 TCAGGGTGACGACAGCTTCCTCAGGGGAGCAGCAGGCCCCCGTCAGGATCTCCAC 442  
Qy 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
Db 441 -----ACCATGACTGTGCAATCTCGGAGCCACTGGCAATGACGTTA 400  
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
Db 399 TCATTG-----TCGGGACACGAGCGATCCAGCAGCGGGCTGTGGCCACAG 349  
Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
Db 348 ACCGTGGCGCATTTGTCACACGCTCCAGTCTCCAGGGGAGCAGCAGGAGGCG 289  
Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgLysSerArgLysIle 94  
Db 288 CTCTCCCGCGTGGCTCACATATAGGGCCACGAATCTGGGGTTGACAGCAG----- 235  
Qy 95 ArgGluMetAsnLysThrIleSerGlnGlu-SerAlaArgValAsnHisArgLeuProGln 114  
Db 234 -----AAGCCACTGTCCCGAGTGTCTCGGAGCGCGCATCTCTATAACAC 187  
Qy 114 uGlyHisProLeuLeu 119  
Db 186 TGGTCGGCCTTGGCTG 171

RESULT 8  
CD467297/c  
LOCUS  
DEFINITION LeukoS1\_3\_A07.gi\_A023 Stimulated peripheral blood leukocytes S1  
Equis caballus cDNA clone LeukoS1\_3\_A07\_A023 5', mRNA sequence.  
ACCESSION CD467297  
VERSION CD467297.1 GI:31388565  
KEYWORDS EST.  
SOURCE Equis caballus (horse)  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Perissodactyla; Equidae; Equis.

REFERENCE 1 (bases 1 to 671)  
Vandenplas, M., Cordonnier-Pratt, M.-M., Suzuki, Y., Sugano, S., Moore, J.N., Liang, C., Sun, F., Sullivan, R., Shah, M. and Pratt, L.H.  
An EST database from equine (Equis caballus) stimulated peripheral blood leukocytes  
Unpublished  
Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu

LIBRARY constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; tissue and RNA were prepared in the Department of Large

Animal Medicine, University of Georgia; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.

Seq primer: Sug5 (CTTCTGCTCTAAAGCTGGG).

FEATURES

Location/Qualifiers

1..671  
/organism="Equis caballus"  
/mol\_type="mRNA"  
/strain="Belgian draft"  
/db\_xref="taxon:9796"  
/clone="LeukoS1\_3\_A07\_A023"  
/sex="female"  
/tissue\_type="blood"  
/cell\_type="leukocytes"  
/lab\_host="DH10B-T1 phage-resistant E. coli"  
/clone\_lib="Stimulated peripheral blood leukocytes S1"  
/note="Organ: circulating peripheral blood leukocytes; Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI. The library was prepared from polyA+ RNA from equine peripheral blood leukocytes isolated from a healthy adult horse. The leukocytes were stimulated for 4 hr with 10 ng/ml E. coli O55:B5 LPS. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5'-prime DraIII site is CACTGTGTG, 3'-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."

BASE COUNT 119 a 200 c 211 g 141 t

ORIGIN

Alignment Scores:  
Pred. No.: 15 Length: 671  
Score: 95.50 Matches: 39  
Percent Similarity: 43.65% Conservative: 16  
Best Local Similarity: 30.95% Mismatches: 47  
Query Match: 13.49% Indels: 24  
DB: 14 Gaps: 5

US-10-087-573-2 (1-141) x CD467297 (1-671)

Qy 3 SerThrSerThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
Db 497 TCCAGGGTGACGACAGCTTCCTCAGGGGAGCAGCAGGCCCCCGTCAGGATCTCCAC 438  
Qy 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
Db 437 -----ACCATGACTGTGCAATCTCGGAGCCACTGGCAATGACGTTA 396  
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
Db 395 TCATTG-----TCGGGACACGAGCGATCCAGCAGCGGGCTGTGTGGCCACAG 345  
Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
Db 344 ACCGTGGCGCATTTGTCACACGCTCCAGTCTTCGCCAGGGGAGCAGCAGGAGGCG 285  
Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
Db 284 CTCTCCCGCGTGGCTCACATATAGGGCCACGAATCTGGGGTTGACAGCAG----- 231  
Qy 95 ArgGluMetAsnLysThrIleSerGlnGlu-SerAlaArgValAsnHisArgLeuProGln 114  
Db 230 -----AAGCCACTGTCCCGAGTGTCTCGGAGCGCGCATCTCTATAACAC 183  
Qy 114 uGlyHisProLeuLeu 119  
Db 182 TGGTCGGCCTTGGCTG 167

RESULT 9  
CD468492/c  
LOCUS  
DEFINITION LeukoS3\_3\_D08.gi\_A025 Stimulated peripheral blood leukocytes S3  
Equis caballus cDNA clone LeukoS3\_3\_D08\_A025 5', mRNA sequence.

ACCESSION CD468492  
 VERSION CD468492.1 GI:31389760  
 KEYWORDS EST.  
 SOURCE Equus caballus (horse)  
 ORGANISM Equus caballus  
 REFERENCE 1 (bases 1 to 700)  
 AUTHORS Vandenplas, M., Cordonnier-Pratt, M.-M., Suzuki, Y., Sugano, S., Moore, J.N., Liang, C., Sun, F., Sullivan, R., Shah, M. and Pratt, L.H.  
 TITLE An EST database from equine (Equus caballus) stimulated peripheral blood leukocytes  
 JOURNAL Unpublished  
 COMMENT Other ESTs: LeukoS3\_3\_D08\_b1\_A025  
 Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; tissue and RNA were prepared in the Department of Large Animal Medicine, University of Georgia; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.  
 Seq primer: Sug5 (CTCTGCTCTAAAGCTGCG).  
 Location/Qualifiers  
 1..700  
 /organism="Equus caballus"  
 /mol\_type="mRNA"  
 /strain="Tennessee walking horse"  
 /db\_xref="taxon:9796"  
 /clone="LeukoS3\_3\_D08\_A025"  
 /sex="female"  
 /tissue\_type="blood"  
 /cell\_type="leukocytes"  
 /lab\_host="DH10B-Ti phage-resistant E. coli"  
 /clone\_lib="Stimulated peripheral blood leukocytes S3"  
 /note="Organ: circulatory system; Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from polyA+ RNA from equine peripheral blood leukocytes isolated from a healthy adult horse. The leukocytes were stimulated for 4 hr with 10 ng/ml E. coli O55:B5 LPS. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."  
 122 a 213 c 217 g 148 t  
 BASE COUNT  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 15.9 Length: 700  
 Score: 95.50 Matches: 39  
 Percent Similarity: 43.65% Conservative: 16  
 Best Local Similarity: 30.95% Mismatches: 47  
 Query Match: 13.43% Indels: 24  
 DB: 14 Gaps: 5  
 US-10-087-573-2 (1-141) x CD468492 (1-700)  
 QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGluThrPhe 22  
 Db TCCAGGTTGACGACGACTTCCTCAGGGGCGAGCACCAGCCCGCTCAGGATCTCCAC 441  
 QY 23 AspValMetArgGluAlaLeuLeuArgValIysSerSerGluArgLeuAlaMetLeuArg 42  
 Db ACCATGACTGTGCAATCTCTCGAGCCCACTGGCAATGACGTTA 399  
 QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58

Db 398 TCATTG-----TCGGGGCACCAGGCGATCCAGCAGCGGGCTGTGTGCCACAG 348  
 QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
 Db 347 ACCGTGGCGGCATTCTTGTCCACACGTCAGTCTTCCAGGGGCGAGCACCAGGAGGCC 288  
 QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
 Db 287 CTCCCGCCGTGGCTCCATATGAGGGGCCACGAACCTTGGGTTGACAGCAGCAG----- 234  
 QY 95 ArgGluMetAsnLysThrIleSerGlnGlu-SerAlaArgValAsnHisArgLeuProGln 114  
 Db 233 -----AAGCCACTGTCCAGTGGTCTGGAGAGCGGCACATCTCTATAACAC 186  
 QY 114 uGlyHisProLeuLeu 119  
 Db 185 TGGTCGCGCTTGGCTG 170  
 RESULT 10  
 LOCUS CD535377/c  
 DEFINITION LeukONS\_2\_B09\_g1\_A027 Unstimulated peripheral blood leukocytes N5  
 Equus caballus cDNA clone LeukONS\_2\_B09\_A027 5', mRNA sequence.  
 ACCESSION CD535377  
 VERSION CD535377.1 GI:31577792  
 KEYWORDS EST.  
 SOURCE Equus caballus (horse)  
 ORGANISM Equus caballus  
 REFERENCE 1 (bases 1 to 729)  
 AUTHORS J.N., Liang, C., Sun, F., Sullivan, R., Shah, M. and Pratt, L.H.  
 TITLE An EST database from equine (Equus caballus) unstimulated peripheral blood leukocytes  
 JOURNAL Unpublished  
 COMMENT Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; tissue and RNA were prepared in the Department of Large Animal Medicine, University of Georgia; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.  
 Seq primer: Sug5 (CTCTGCTCTAAAGCTGCG).  
 Location/Qualifiers  
 1..729  
 /organism="Equus caballus"  
 /mol\_type="mRNA"  
 /strain="Dartmoor Pony"  
 /db\_xref="taxon:9796"  
 /clone="LeukONS\_2\_B09\_A027"  
 /sex="male"  
 /tissue\_type="blood"  
 /cell\_type="leukocytes"  
 /lab\_host="DH10B-Ti phage-resistant E. coli"  
 /clone\_lib="Unstimulated peripheral blood leukocytes N5"  
 /note="Organ: circulatory system; Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from polyA+ RNA from unstimulated equine peripheral blood leukocytes isolated from a healthy adult horse. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."  
 FEATURES  
 source



REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	(bases 1 to 881)	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:sgapbs-remail.nih.gov">sgapbs-remail.nih.gov</a> Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLAM11367 row: f column: 19 High quality sequence stop: 877. Location/Qualifiers 1. .881 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="IMAGE:5148690" /tissue_type="tumor, gross tissue" /dev_stage="7 months" /lab_host="DH10B" /clone_lib="NCI CGAP Mam5" /note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NIH"
BASE COUNT	168 a	251 c	279 g	183 t
ORIGIN				
Alignment Scores:				

```

BASE COUNT      168 a  251 c  279 g  183 t
ORIGIN

Alignment Scores:
Pred. No.:      23.8
Score:          95.00
Percent Similarity: 38.06%
Best Local Similarity: 27.74%
Query Match:    13.42%
DB:             12
                Length: 881
                Matches: 43
                Conserved: 16
                Mismatches: 36
                Indels: 60
                Gaps: 9

US-10-087-573-2 (1-141) x BI248636 (1-881)
QY      3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22
Db      405 TCAAGGTGACGACAGGCTCCGCAAGGGCAGCACCAGGCCCATCCGGATCTCCAC 346
QY      23 AspValMetArgGluAlaLeuArgValIysSerSerGluArgLeuAlaMetLeuArg 42
Db      345 ACCATA-----ACTGTGCAGTCTCAGAGCCACTGGCAATGACATTG 304
QY      43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58
Db      303 TCATTG-----TGTGGACACCGGAGTGTACAGCGGGCAGTGTGGCGCGAG 253
QY      59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74
Db      252 ACCAGGGGCAGCTTCTGTCTACTCGTCCAGTCTTGCTAGGGGTAGCACGAGAGGCC 193
QY      75 LysProProArgProGlnSerThrLysSerProGluLeuArgGlnLeuSerArgLysIle 94
Db      192 CTTCCCCCA-----CTGGCTCACAGATC 169
QY      95 ArgGluMetAsn-----LysThrIleSerGln----- 103
Db      168 AGAGCCATGAACCTGGGGTTGACAGCGCAGAGCCACTGTCCCAAGTGGTTGTGAGAG 109
QY      104 -----GluSer 105
Db      108 CGCACATCTCATAGCACTGGTTCAGCCTTGCTGGCTGTGTCCAAACACGTGGCGGAATTT 49

```

Alignment Scores: 18.6 Length: 607  
 Score: 94.00 Matches: 42  
 Percent Similarity: 40.74% Conservative: 13  
 Best Local Similarity: 31.11% Mismatches: 35  
 Query Match: 13.28% Indels: 45  
 DBs: 10 Gaps: 7

US-10-087-573-2 (1-141) x BG180114 (1-607)

QY	8	ThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArgGlu	27
DB	511	ACACGCTTTGGTGGCCCTCCAGGGTGACGAGGGCTCC	467
QY	28	AlaLeuLeuArg	37
DB	466	GCACGATCAGGCCCCCATCCGGGATCTCCACACCATGACTGTGCAGTCTCGAGGCCA	407
QY	38	LeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGly	57
DB	406	CTGGCAATGACGTTGTTCATTG	356
QY	58	Ala	69
DB	355	GCTGTGGCCACAGACCGTGGCGGATCTTGTCCACACGTCAGTCTTGCCAGGGGC	296
QY	70	AlaSerMetIysLeuIysProProArgProGlnSerThrIysSerProGluLeuArgGlu	89
DB	295	AGCACCAAGGAGGCCCTCCCGC	272
QY	90	LeuSerArgIysIleArgGluMetAsn	103
DB	271	CTGGCCTCACAGATCAGGCGCCACAACTAGGTTGACAGACAGAGCCACTGTGCCAG	212
QY	104	Glu-SerAlaArgValAsnHisArgLeuProGluGlyHisPro	117
DB	211	GTGTCTGTGAGACGCGACATCTTCATAGCACTGTGCGCCT	169

BI818617 669 bp mRNA linear EST 04-OCT-2000  
 603033383F1 NIH\_MGC\_115 Homo sapiens cDNA clone IMAGE:5174421 5'  
 mRNA sequence.  
 BI818617  
 BI818617.1 GI:15929929  
 EST.  
 SOURCE  
 Homo sapiens (human)  
 Organism  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 669)  
 NIH\_MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-x@mail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L14M11434 row: f column: 22  
 High quality sequence stop: 668.  
 Location/Qualifiers  
 1..669  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5174421"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH MGC 115"

FEATURES  
 source

/note="Organ: pooled brain, lung, testis; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH MGC Library."

BASE COUNT 114 a 211 c 212 g 132 t

Alignment Scores:  
Pred. No.: 21 Length: 669  
Score: 94.00 Matches: 41  
Percent Similarity: 41.98% Conservative: 14  
Best Local Similarity: 31.30% Mismatches: 41  
Query Match: 13.28% Indels: 35  
DB: 12 Gaps: 6

US-10-087-573-2 (1-141) x B1818617 (1-569)

Qy 8 ThrAsnPhValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg--- 26

Db 486 ACACGCTTGGTGGCCCTCCAGGGTGACAGCGGCTCCCGCAGGGGCGAGCATCAGGCCC 427

Qy 27 -----GluAlaLeuArgValLysSerSerGluArgGluAlaMetLeuArg 42

Db 426 CCATCCGGGATCCACACCATGACTGTGCAGTCTCGAGCCACTGGCAATGACGTTG 367

Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58

Db 366 TCATTG-----TGGGGCACCAGCGGATGTAGCACAGGGGCTGTGTGGCCACAG 316

Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74

Db 315 ACCGTGGGGCGCATTTGTCCACAGCTCCAGCTTGGCCAGGGGCGAGCAGGAGGCC 256

Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94

Db 255 CTTCCCGG-----CTGGCCTCACAGATC 232

Qy 95 ArgGluMetAsn-----LysThrIleSerGlnGlu-SerAlaArgVa 108

Db 231 AGGGCCACAACCTTAGGGTTGACAGCACAGAGCCACTGTCCAGGCTGTGTGAGAGC 172

Qy 108 lAsnHisArgLeuProGluGlyHisProLeu 118

Db 171 CGCACATCTTCATAGCAGCTGTGGCCCTTG 141

RESULT 16

CC366660 927 bp DNA linear GSS 16-MAY-2003

LOCUS PUH8B80TD ZM.0.6.1.0 KB Zea mays genomic clone ZM8BTA521M16,

DEFINITION genomic survey sequence.

ACCESSION CC366660

VERSION CC366660.1 GI:30836060

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

REFERENCE 1 (bases 1 to 927)

AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

TITLE Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

JOURNAL clade: Panicoideae; Andropogoneae; Zea.

COMMENT Contact: Cathy Whitelaw

TIGR TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

FEATURES

source

Location/Qualifiers

1..927

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZM8BTA521M16"

/clone\_lib="ZM.0.6.1.0 KB"

/notes="vector: pCR4-TOPO; Site 1: EcoRI, 0.6-1.0 kb high

COT selected genomic DNA library"

BASE COUNT 231 a 286 c 201 g 229 t

ORIGIN

Alignment Scores:

Pred. No.: 31.9 Length: 927

Score: 94.00 Matches: 30

Percent Similarity: 47.19% Conservative: 12

Best Local Similarity: 33.71% Mismatches: 41

Query Match: 13.28% Indels: 6

DB: 29 Gaps: 2

US-10-087-573-2 (1-141) x CC366660 (1-927)

Qy 52 ValLeuProGlyThrGlyAlaSerAlaIleAlaAlaThrValThrProLysGlyAlaSer 71

Db 16 GTTCTCTCAGCTACCCGCTGCGGCCCTTGCTGCGGCTTGCAGCCCTTGCCATCGTCA 75

Qy 72 MetLys-----LeuLysProProArgProGlnSerThrLysSerProGluLeu 87

Db 76 TTGTCTACTGTGCGCTTGTGTCACACACCGCTACCGACATGTGCTTCCGCCCTTG 135

Qy 88 ArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArg 107

Db 136 TCACCACCATCA-----TACTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 189

Qy 108 ValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGluTyrPheArg 127

Db 190 CCACCGCTCGATGCCGCCAGAGGTGGCCCTCTCGGCGCTGATAGCCCATGGAGTCGT 249

Qy 128 HisLeuArgSerLeuLysSerGlnGly 136

Db 250 CCATGCAACGAGCAAGACGAAAGGG 276

RESULT 17

BM909493/c

LOCUS BM909493 1172 bp mRNA linear EST 12-MAR-2002

DEFINITION AGENCOURT\_6640673 NIH\_MGC\_99 Homo sapiens cDNA clone IMAGE:5434084

S', mRNA sequence.

ACCESSION BM909493

VERSION BM909493.1 GI:19359872

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1172)

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabbe-remail.nih.gov

Tissue Procurement: Lou Staudt

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCMI910 row: j column: 05

High quality sequence stop: 515.

FEATURES		REFERENCE	
Source		AUTHORS	
1. 1172		1. (bases 1 to 660)	
/organism="Homo sapiens"		Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nixaldo, I., Osato, N., Saito, R., Suzukhi, H., Yamakawa, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojbori, T., Baidarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Bruscia, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Glissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konegaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Perle, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ranachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semplic, C.A., Secou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wyrshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.	
/db_xref="taxon:9606"		Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	
/clone="IMAGE:5434084"		Nature 420, 563-573 (2002)	
/tissue="lymphoma, cell line"		22354583	
/lab_host="DH10B (phage-resistant)"		12466851	
/note="Organ: lymph. Vector: pOTB7; Site: 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACAGG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."		Contact: Yoshihide Hayashizaki	
BASE COUNT		Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute	
ORIGIN		The Institute of Physical and Chemical Research (RIKEN)	
169 a 463 c 260 g 280 t		1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan	
Alignment Scores:		Tel: 81-45-503-9222	
Pred. No.: 43.1		Fax: 81-45-503-3216	
Score: 94.00		Email: genome-res@gsr.riken.go.jp,	
Percent Similarity: 41.79%		URL: http://genome.gsc.riken.go.jp/	
Best Local Similarity: 30.60%		Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission	
Query Match: 12		Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome 12, 673-677 (2001)	
DB: 12		Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)	
US-10-087-573-2 (1-141) x BM909493 (1-1172)		RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichannel sequencer. Genome Res. 10 (11), 1757-1771 (2000)	
QY 4 ThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAsp 23		Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)	
Db 513 ACCACATGCCACAGCTTGTTGGCTCCAGGGTGACAGCGGCTCCCGAGGGGC 454		cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.	
QY 24 ValMetArg-----GluAlaLeuLeuArgValHisSerSerGluArgLeu 38		Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC Building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.	
Db 453 AGCATCAGGCGCCCATCCGGGATCTCCACACCATGCTGTGCAGTCTCCGAGGCACGTG 394		Please visit our web site ( <a href="http://genome.gsc.riken.go.jp">http://genome.gsc.riken.go.jp</a> ) for further details.	
QY 39 AlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla 58		Location/Qualifiers	
Db 393 GCAATGACGTGTCATG-----TGGGGCACCGACGGCTGCTGACAGCGGGCT 343		1. .660	
QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAla 70		FEATURES	
Db 342 GTGTGGCCACAGACCGTGGGGCGCATTTCTTCCACACAGTCCAGTCTTGGCCAGGGGCAGC 283		source	
QY 71 SerMetLysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeu 90		BY749167 660 bp mRNA linear EST 17-DEC-2002	
Db 282 ACCAGGAAGCCCTCCCGG-----LysThrIleSerGlnGlu 104		BY749167 RIKEN full-length enriched, NOD-derived cDNA +ve	
QY 91 SerArgGlyIleArgGluMetAsn-----LysThrIleSerGlnGlu 104		dendritic cells Mus musculus cDNA clone F630230F14 5', mRNA	
Db 258 GCCTCAGATCAGGCGCCACAACTTAGGTTGACAGCACAGAGCCATGTCACAGGTG 199		sequence.	
QY 105 -SerAlaArgValAsnHisArgLeuProGluGlyHisPro 117		BY749167	
Db 198 GTCGTGAGACGGCGCATCTTTCATAGCATGCTGGCGCT 159		BY749167	
RESULT 18		VERSION	
BY749167/c		KEYWORDS	
LOCUS		SOURCE	
DEFINITION		ORGANISM	
ACCESSION		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
VERSION		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
KEYWORDS			
SOURCE			
ORGANISM			



```

source
1..554
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="EIOV018D12"
/tissue_type="ovary"
/dev_stage="fetus"
/lab_host="DH10B"
/clone_lib="Bos taurus ovary fetus"
/notes="Vector: pZLL; Site_1: SalI; Site_2: NotI; Poly A
was deleted from a NotI site"
BASE COUNT 165 a 135 c 149 g 105 t
ORIGIN

Alignment Scores:
Pred. No.: 23.2 Length: 554
Score: 92.50 Matches: 35
Percent Similarity: 49.59% Conservatve: 25
Best Local Similarity: 28.93% Mismatches: 45
Query Match: 13.06% Indels: 16
DB: 9 Gaps: 5

US-10-087-573-2 (1-141) x AV667883 (1-554)

QY 26 ArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArgAlaLeuAla 45
|||||
Db 40 CGGGATCTCTCCCTCCGGTTGAGACGAGTGAATATGCTTCTGTGGGATTGGAG 99
|||||

QY 46 GlyMetCysGlyHisArgValLeuPro-GlyThrGlyValSerAlaAlaAlaThrVa 65
|||||
Db 100 -----TGTGGATAAACAAGATCCCTGGGCATCGGGAATCGCTCTCCGGGAGTGA 153
|||||

QY 65 IThrProLysGlyAlaSerMetLysLeuLysProProArgPro-----G1 80
|||||
Db 154 GAGGAGCGCATAAAAGAAATGCCCCGGTGAAGCCACAGACACCGCTCCTACAGAGCAG 213
|||||

QY 80 rSerThrLysSerProGluLeuArgGlu---LeuSerArgLysIleArgGluMetAsnLys 99
|||||
Db 214 GAGTTGCCACAGCCCCAGGCTGACACAGGCTCTGGACAGATCTGATGATGATGAATCA 273
|||||

QY 99 sThrIleSerGlnGlnSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLe 119
|||||
Db 274 GTCCACAGAGCTTGAGGAAC---AGGATTCTACACAGGCAACCA-----313
|||||

QY 119 uGluLysArgAlaGluThrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnAr 139
|||||
Db 314 -CACACACAGCTCAGCTGGCAGCAGCAGCTGAATCGATGAGAACACCTCAGTAAAGCA 372
|||||

QY 139 g 139
Db 373 A 373

RESULT 20
AV597776
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE

```

PUBMED 11713328  
 COMMENT Contact: Yoshikazu Sugimoto  
 Animal Genetics Division  
 Shirakawa Institute of Animal Genetics  
 Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan  
 Tel: 81-248-25-5641  
 Fax: 81-248-25-5725  
 Email: kazuugi@coo.ocn.ne.jp  
 Single pass sequencing  
 This clone was obtained from a polyA-deleted cDNA library.  
 Location/Qualifiers  
 1..578  
 /organism="Bos taurus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9913"  
 /clone="E1CA034A05"  
 /tissue\_type="cartilage"  
 /dev\_stage="fetus"  
 /lab\_host="DH10B"  
 /clone\_lib="Bos taurus cartilage fetus"  
 /note="Vector: pZ11; Site\_1: SalI; Site\_2: NotI; Poly A was deleted from a NotI site"

BASE COUNT 163 a 147 c 173 g 95 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 24.5 Length: 578  
 Score: 92.50 Matches: 35  
 Percent Similarity: 49.59% Conservative: 25  
 Best Local Similarity: 28.93% Mismatches: 45  
 Query Match: 13.06% Indels: 16  
 DB: 9 Gaps: 5  
 US-10-087-573-2 (1-141) x AV597776 (1-578)

Qy 26 ArgGluAlaLeuLeuArgValLysSerGluArgLeuAlaMetLeuArgAlaLeuAla 45  
 Db 144 CGGATCCTCTCCCTCGGTTGAGAACGAGTAAATGCTTCGTGCGGATTGGAG 203  
 Qy 46 GlyMetCysGlyHisArgValLeuPro-GlyThrGlyAlaSerAlaAlaAlaThrVa 65  
 Db 204 -----TGTTGAATAAACGAATCCCTGGGCACTGGGAAATCCGCTCTCCCGGAGTGAA 257  
 Qy 65 lThrProLysGlyAlaSerMetLysLeuLysProArgPro-----G1 80  
 Db 258 GAGGAGCGATAAAAGAAATCCCGTGAAGCCACAGACACCGTCCCTGTACAGAGCAG 317  
 Qy 80 nSerThrLysSerProGluLeuArgGlu---LeuSerArgLysIleArgGluMetAsnLy 99  
 Db 318 GAGTTGCCACACCCAGGCTGAGACAGGCTCTGGAACAGAAATCTGATAGTGAATCA 377  
 Qy 99 sThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLe 119  
 Db 378 GTCCAGAGCTTTGAGGAAC--AGGATTCTACAGGCAACCA-----417  
 Qy 119 uGluLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnAr 139  
 Db 418 -CACACAAGCTCAGCTGGCAGCAGCAGCTGAATCGATGAAGAACCACTCAGTAAAGCA 476  
 Qy 139 g 139  
 Db 477 A 477

RESULT 21  
 AV616040  
 LOCUS  
 DEFINITION AV616040 Bos taurus ovary fetus Bos taurus cDNA clone E1OV002H02  
 5', mRNA sequence.  
 ACCESSION AV616040  
 VERSION AV616040.1 GI:9751710  
 KEYWORDS EST.  
 SOURCE Bos taurus (cow)  
 ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovidae; Bovinae; Bos.  
 1 (bases 1 to 578)  
 Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.  
 and Sugimoto,Y.  
 Establishment of a high throughput EST sequencing system using  
 poly(A) tail-removed cDNA libraries and determination of 36,000  
 bovine ESTs  
 Nucleic Acids Res. 29 (22), E108 (2001)  
 21570554  
 11713328  
 COMMENT Contact: Yoshikazu Sugimoto  
 Animal Genetics Division  
 Shirakawa Institute of Animal Genetics  
 Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan  
 Tel: 81-248-25-5641  
 Fax: 81-248-25-5725  
 Email: kazuugi@coo.ocn.ne.jp  
 Single pass sequencing.  
 This clone was obtained from a polyA-deleted cDNA library.  
 Location/Qualifiers  
 1..578  
 /organism="Bos taurus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9913"  
 /clone="E1OV002H02"  
 /tissue\_type="ovary"  
 /dev\_stage="fetus"  
 /lab\_host="DH10B"  
 /clone\_lib="Bos taurus ovary fetus"  
 /note="Vector: pZ11; Site\_1: SalI; Site\_2: NotI; Poly A was deleted from a NotI site"

BASE COUNT 168 a 142 c 156 g 111 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 24.5 Length: 578  
 Score: 92.50 Matches: 35  
 Percent Similarity: 49.59% Conservative: 25  
 Best Local Similarity: 28.93% Mismatches: 45  
 Query Match: 13.06% Indels: 16  
 DB: 9 Gaps: 5  
 US-10-087-573-2 (1-141) x AV616040 (1-578)

Qy 26 ArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArgAlaLeuAla 45  
 Db 58 CGGATCCTCTCCCTCGGTTGAGAACGAGTAAATGCTTCGTGCGGATTGGAG 117  
 Qy 46 GlyMetCysGlyHisArgValLeuPro-GlyThrGlyAlaSerAlaAlaAlaThrVa 65  
 Db 118 -----TGTTGAATAAACGAATCCCTGGGCACTGGGAAATCCGCTCTCCCGGAGTGAA 171  
 Qy 65 lThrProLysGlyAlaSerMetLysLeuLysProArgPro-----G1 80  
 Db 172 GAGGAGCGATAAAAGAAATGCGCGTGAAGCCACAGACACCGTCCCTGTACAGAGCAG 231  
 Qy 80 nSerThrLysSerProGluLeuArgGlu---LeuSerArgLysIleArgGluMetAsnLy 99  
 Db 232 GAGTTGCCACACCCAGGCTGAGACAGGCTCTGGAACAGAAATCTGATAGTGAATCA 291  
 Qy 99 sThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLe 119  
 Db 292 GTCCAGAGCTTGAAGAAC--AGGATTCTACAGGCAACCA-----331  
 Qy 119 uGluLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnAr 139  
 Db 332 -CACACAAGCTCAGCTGGCAGCAGCAGCTGAATCGATGAAGAACCACTCAGTAAAGCA 390  
 Qy 139 g 139  
 Db 391 A 391

```

RESULT 22
AV601755      594 bp   mRNA   linear   EST 27-NOV-2001
LOCUS       AV601755 Bos taurus kidney fetus Bos taurus cDNA clone E1KI006B10
DEFINITION   5', mRNA sequence.
ACCESSION   AV601755
VERSION     AV601755.1 GI:9724077
SOURCE      EST
ORGANISM    Bos taurus (cow)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 594)
Takaasuga,A., Hirotsune,S., Itoh,R., Jitchazono,A., Suzuki,H., Aso,H.
and Sugimoto,Y.
Establishment of a high throughput EST sequencing system using
poly(A) tail-removed cDNA libraries and determination of 36,000
bovine ESTs
Nucleic Acids Res. 29 (22), E108 (2001)
21570554
11713328
Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-Shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@cocoa.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.
FEATURES             source
BASE COUNT          170 a 152 c 160 g 112 t
ORIGIN
...
Alignment Scores:
Pred. No.:           25.4               Length:           594
Score:              92.50               Matches:           35
Percent Similarity: 49.59%              Conservative:      25
Best Local Similarity: 28.93%            Mismatches:        45
Query Match:         13.06%              Indels:           16
DB:                  Gaps:                5

US-10-087-573-2 (1-141) x AV601755 (1-594)

QY      26 ArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArgAlaLeuAla 45
Dbbbbb:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      73 CGGGATCCTCTCCCTCGGGTTGAAGCCGAGTCAGTAATAATGCTCTGTGTCGGATTGGAG 132
QY      46 GlyMetCysGlyHisargValLeuPro-GlyThrGlyAlaSerAlaIleAlaAlaThrVa 65
Dbb-----TGTTGGATAAAACGAATCCTTGGGCATCGGAATTCGCCTCCGCCGAGTGAA 186
QY      65 lThrProLysGlyAlaSerMetLysLeuLysProProArgPro-----GI 80
Db      187 GAGGAGCGATAAAAAAATAATGCCGGTTGAAGCCACAGACACCCTCCCTGCTACAGACGAG 246
QY      80 nSerThrLysSerProGluLeuArgGlu---LeuSerArgLysIleArgGluMetAsnLy 99
Db      247 GAGTTGCCACAGCCCAGCGCTGAGACAGGGTCITGGAACAGAATCTGATAGTGAATCA 306

```

```

DB:
US-10-087-573-2 (1-141) x BF371432 (1-615)
QY 3 SerThrSerThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22
Db 141 GCAAGCACCACTCCCAAGGAGGTGCACCGCCCTACTAATAGC----- 191
QY 23 AspValMetArgGluAlaLeuLeuArgValLysSerGluArgLeuAlaMetLeuArg 42
Db 192 -----AGGTTGAGTTCAGTGCACCGCCCTACTAATAGC----- 227
QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeu-----ProGlyThrGlyAlaSerAla 60
Db 228 GTGATATCCAGGTGCACGACCTGTGGCACTCCGGAAACAGGAGAGCCCGCTGTCT 287
QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProArgProGln 80
Db 288 ATCACCCTCCAGCTCAGTCCGCTGTGTTCCTCCGCGGCTCCGCGCCACCCCGAG 347
QY 81 ---SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArg----- 95
Db 348 GTGGCTGTTCGCTGCTGTGGCCGACTGTCACCGCAGGTACAGGCTCTGTAGTC 407
QY 96 -----GluMetAsnLysThrIleSerGlnGluSerAlaArgVal-AsnHisAr 111
Db 408 TGGTTAATACCAAAATAGGGGAAGAGGTGGCAAGGAA---TACAGAACCGCCATCG 464
QY 111 gLeuProGluGlyHis-----ProLeuLeuGluLysArgAlaGluTyrPheArgHis 129
Db 465 GCACCCCATCTCTGTGGCTGTGACGTGACCAAGCTGTGGCCACCGCGGACCT 524
QY 129 uArgSerLeuLysSer 134
Db 525 TCGGGCACTTGAAGT 540

RESULT 24
BES74311
LOCUS
DEFINITION
SNEST4a12c04.y1 csn 1 S neurona invitro merozoite cDNA Sarcocystis
neutona cDNA 5, similar to TR:Q13765 Q13765 NASCENT POLYPEPTIDE
ASSOCIATED COMPLEX ALPHA SUBUNIT. ;, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
On Aug 14, 2000 this sequence version replaced gi:9824355.
Contact: Daniel K. Howe
Sarcocystis neurona EST project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Contact Daniel K. Howe (dkhowe2@pop.uky.edu) for further
information relating to organism, libraries, or clone availability.
Seq primer: -40RP from GIBCO
High quality sequence stop: 426.
Location/Qualifiers
1. .641
/organism="Sarcocystis neurona"

/mol_type="mRNA"
/strain="Gn3"
/db_xref="taxon:42890"
/dev_stage="merozoite"
/lab_hosts="DH10B"
/clone_lib="csn 1 S neurona invitro merozoite cDNA"
note=vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; The library was constructed by Dan Howe, University
of Kentucky. cDNAs were synthesized from poly(A)+ RNA
by oligo d(T) priming and directionally cloned into the
Uni-ZAP XR lambda vector. The library was mass excised
as phagemids and rescued in SOLR cells. The plasmid
library was recovered from the SOLR cells and transformed
in mass into DH10B cells for sequencing. WARNING: the
library contains a small percentage of cDNAs derived from
the bovine host cells."
BASE COUNT 203 a 149 c 167 g 120 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 28 Length: 641
Score: 92.50 Matches: 35
Percent Similarity: 49.59% Conservative: 25
Best Local Similarity: 28.93% Mismatches: 45
Query Match: 13.06% Indels: 16
DB: 10 Gaps: 5
US-10-087-573-2 (1-141) x BES74311 (1-641)
QY 26 ArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArgAlaLeuAla 45
Db 21 CGGGATCTCTCCCTCGGGTTGAGAACCGCATTAATGCTTCTGTGGGATTTGGAG 80
QY 46 GlyMetCysGlyHisArgValLeuPro-GlyThrGlyAlaSerAlaIleAlaThrVa 65
Db 81 -----TGTGAATAAAGCAATCCCTGGGCACTGGGAAATCCGCTCTCCCGGAGTGAA 134
QY 65 lThrProLysGlyAlaSerMetLysLeuLysProArgPro-----G1 80
Db 135 GAGGAGCGCAATAAAGAAATCCCGGTGAAGCCACAGACCCGCTCTGTCTACAGAGCAG 194
QY 80 rSerThrLysSerProGluLeuArgGlu-----LeuSerArgLysIleArgGluMetAsnLy 99
Db 195 GAGTTGCCACAGCCCGGCTGAGACAGGGTCTGGACAGCAATCTGATAGTGAATCA 254
QY 99 sThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLe 119
Db 255 GTCCACAGAGTTGAGGAAC---AGGATTTCTACACAGGCAACCA----- 294
QY 119 uGluLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnAr 139
Db 295 -CACAAAGCTCAGCTGGCAGCAGCTGAAATCGATGAAGAACCAAGTCAGTAAAGCA 353
QY 139 g 139
Db 354 A 354

RESULT 25
BPI12124
LOCUS
DEFINITION
BPI12124 ORCS bovine utero-placenta cDNA Bos taurus cdna clone
ORCS12748 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Bos taurus (cow)
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 667)
Ishiwata,H., Katsuma,S., Kizaki,K., Patel,O.V., Nakano,H.,
Takahashi,T., Imai,K., Hirasawa,A., Shiojima,S., Ikawa,H., Suzuki

```



```

Db      201 ACAGCACAGAGGCACTGCTCCAGAGTGCTGTGTGAGAGCGGACACATCTTCATAGCACTGG 142
Qy      115 YHisProLeu 118
          |||
Db      141 TCGGCCCTTG 132

RESULT 27
BI757298/c
LOCUS   603029512F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5199795 5',
DEFINITION
ACCESSION BI757298
VERSION   BI757298
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          1 (bases 1 to 875)
          NIH-MGC http://mgc.nci.nih.gov/.
          National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished
          Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Life Technologies, Inc.
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM1500 row: h column: 04
          High quality sequence stop: 793.
          Location/Qualifiers
            1..875
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:519795"
              /lab_host="DH10B"
              /clone_lib="NIH_MGC_114"
              /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
              Site 2: EcoRV (destroyed); RNA source anonymous pool of 6
              male brains, age range 23-27 yo. Library is oligo-dT
              primed and directionally cloned (EcoRV site is destroyed
              upon cloning). Average insert size 1.5 kb, insert size
              range 1-3 kb. Library is normalized and enriched for
              full-length clones and was constructed by C. Gruber
              (Invitrogen). Research Genetics tracking code 019. Note:
              this is a NIH MGC Library."
              this is a NIH MGC Library."
BASE COUNT 156 a 272 c 277 g 170 t
ORIGIN
Alignment Scores:
Pred. No.: 41.7 Length: 875
Score: 92.50 Matches: 38
Percent Similarity: 46.58% Conservative: 30
Best Local Similarity: 26.03% Mismatches: 52
Query Match: 13.06% Indels: 26
DB: 12 Gaps: 6

US-10-087-573-2 (1-141) x BI757298 (1-875)

Qy      3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22
          |||
Db      409 GCAAGCACCAACACAACTCCCAAGGAGGCTGCCAGGCGCCCTACTAATAGC----- 359
Qy      23 AspValMetArgGluAlaLeuLeuArgVallySerSerGluArgLeuAlaMetLeuArg 42
          |||
Db      358 -----AGGTTGAGGTGAGTGCAGTGCAGGCCCGGCGGTGTGG 323
Qy      43 AlaLeuAlaGlyMetCysGlyHisArgValLeu-----ProGlyThrGlyAlaSerAla 60
          |||

```

```

Db      322 GTGATATCCACGGTCAACAGCACTGGGCACCTCCGGGAACACGAGGAGGCCCCCGGTGCT 263
Qy      61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProArgProGln 80
          |||
Db      262 ATCACCTCCAGCTCAGCTCCCGTGTGCTTCCCGCGGGTCCGGCCACCCAGAG 203
Qy      81 ---SerThrLysSerProGluLeuArgGluLeuSerArgLysLleArg----- 95
          |||
Db      202 GTGCGTCTCCCGTGTGCTGCCGACCTGTCCACCCGAGGTACAGGGCTCTCTGTAGTC 143
Qy      96 -----GluMetAsnLysThrIleSerGlnGluSerAlaArgVal-AsnHisAr 111
          |||
Db      142 TGGTTAATACCAAAATAGGGGAGAGGTGGCAGGGA---TACAGAACCAGGCCATCG 86
Qy      111 gLeuProGluGlyHis-----ProLeuLeuGluLysArgAlaGluThrPheArgHisLe 129
          |||
Db      85 GCACCCCATCTCTCATCTGTGGCTGTGCTGACCCCAAGCTGTGGCCAGCGGCACCT 26
Qy      129 uArgSerLeuLysSer 134
          |||
Db      25 TCGGGCACITGGAGT 10

RESULT 28
BF301271/c
LOCUS   602029713F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4164939 5',
DEFINITION
ACCESSION BF301271
VERSION   BF301271
KEYWORDS  EST.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
          1 (bases 1 to 907)
          NIH-MGC http://mgc.nci.nih.gov/.
          National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished
          Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Jeffrey E. Green, M.D.
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM9451 row: i column: 04
          High quality sequence stop: 686.
          Location/Qualifiers
            1..907
              /organism="Mus musculus"
              /mol_type="mRNA"
              /strain="FVB/N"
              /db_xref="taxon:10090"
              /clone="IMAGE:4164939"
              /lab_host="DH10B (71 phage-resistant)"
              /clone_lib="NCI_CGAP_SG2"
              /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:
              NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo
              dT. Average insert size 1.3 kb. Constructed by Life
              Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 139 a 278 c 299 g 191 t
ORIGIN
Alignment Scores:
Pred. No.: 43.7 Length: 907
Score: 92.50 Matches: 40
Percent Similarity: 41.22% Conservative: 14
Best Local Similarity: 30.53% Mismatches: 41
Query Match: 13.06% Indels: 36
DB: 10 Gaps: 6

```

Carlis, E.D., Kanai, A., Kawaj, H., Kawasawa, Y., Kedzierski, K.M.,

```

Score:          91.50      Matches:          30
Percent Similarity: 46.59%      Conservative:    11
Best Local Similarity: 34.09%      Gaps:            5
Query Match:      12.92%      Indels:          29
DB:               13          Gaps:            5

US-10-087-573-2 (1-141) x BY224833 (1-425)

Qy 30 LeuArgValLysSerSerGluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGly 49
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 421 ATAACTGTGCAGTCTCTCAGAGCCACTGGCAATGACATTTCATTG-----TGTGGA 371
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

Qy 50 HisArgValLeuProGlyThrGlyAla-----SerAlaIleAlaAla 63
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 370 CACGAGCGCATGTCTAGCAGAGGGGCAGTGTGCCCGCAGACGAGGGGCACGTTCTTGTCT 311
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

Qy 64 Thr-----ValThrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSer 81
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 310 ACTCGTCCAGTCTGCTAGGGGTAGCAGAGAGGCCCTCCCCCA-----263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 82 ThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsn-----98
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 262 -----CTGGCGCTCACAGATCAGAGCCATGAACTTGGGGTTG 227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 99 -----LysThrIleSerGln 103
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 226 ACAGCGCAGAGCCACTGTCCCAA 203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 31  
BE553346/c  
LOCUS  
DEFINITION  
          BE553346 linear EST 15-AUG-2000  
          ur45D04.y1 NCI\_CGAP Mam2 Mus musculus cDNA clone IMAGE:3153199 5'  
          similar to SW:CORO\_BOVIN Q92176 CORONIN-LIKE PROTEIN P57. ;, MRNA  
          sequence.  
ACCESSION  
          BE553346  
VERSION  
          BE553346.1 GI:9817833  
KEYWORDS  
          EST.  
SOURCE  
          Mus musculus (house mouse)  
ORGANISM  
          Mus musculus  
REFERENCE  
          Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS  
          1 (bases 1 to 544)  
TITLE  
          NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
          National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
          Tumor Gene Index  
JOURNAL  
          Unpublished  
COMMENT  
          Contact: Robert Strausberg, Ph.D.  
          Email: [cgapsb@mail.nih.gov](mailto:cgapsb@mail.nih.gov)  
          This clone is available royalty-free through LLNL ; contact the  
          IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

```

Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgyabs@emall.nih.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:1055955
Seq primer: -40RP from Gibco
High quality sequence stop: 363.
Location/Qualifiers
1..544
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:3153199"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP Mam2"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
101 a 176 q 11 t 2 others
BASE COUNT
101 a 154 c

```

```

/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:3153199"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI CGAP Mam2"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Öligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT      101 a      154 c      176 g      111 t      2 others
ORIGIN
Alignment Scores:      28.5      Length:      544
Pred. No.:

```

ORIGIN	Alignment Scores:	
	Pred. No.:	Length:
	28.5	544



Score: 91.50 Matches: 35  
Percent Similarity: 42.61% Conservative: 14  
Best Local Similarity: 30.43% Mismatches: 31  
Query Match: 12.92% Indels: 35  
DB: 10 Gaps: 6

US-10-087-573-2 (1-141) x BE553346 (1-544)

QY 3 SerThrSerThrThrThrAsnPhValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
DB 357 TCCAGGTGACGACAGGCTCCCGAAGGCGACGACCCGCCATCCGGATCTCCAC 298  
QY 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
DB 297 ACCATA-----ACTGTGCAGTCTCTCAGAGCCACTGGCAATGACATTG 256  
QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
DB 255 TCATTG-----TGTGGACACAGCGGATCTAGCAGAGGCGAGTGTGCCCGCAG 205  
QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
DB 204 ACCAGGGCGACGTTCTTGTCTACTCGTCCAGTCTTGCCTAGGGTAGCACCAGGAAGGCC 145  
QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
DB 144 COTCCCCCA-----CTGGCCTCAGATC 121  
QY 95 ArgGluMetAsn-----LysThrIleSerGln 103  
DB 120 AGAGCCATGAAGTGGGGTTGACAGCGCAGAGGACCACTGTCCCAA 76

RESULT 32  
BE284887/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BE284887 640 bp mRNA linear EST 26-OCT-2000  
601097563F1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:3496502 5',  
mRNA sequence.  
BE284887 BE284887.1 GI:9162136  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 640)  
NIH-MGC <http://mgi.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-rc@mail.nih.gov](mailto:cgabs-rc@mail.nih.gov)  
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM8548 row: m column: 15  
High quality sequence stop: 612.  
Location/Qualifiers  
1..640  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3496502"  
/tissue\_type="tumor, gross tissue"  
/dev\_stages="7 months"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP\_Mam5"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: Sali;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies. Investigators

FEATURES  
source

providing samples: Lothar Hennighausen/Robin Humphreys,  
NIH"

BASE COUNT 114 a 188 c 202 g 136 t  
ORIGIN

Alignment Scores:  
Pred. No.: 35.1 Length: 640  
Score: 91.50 Matches: 35  
Percent Similarity: 42.61% Conservative: 14  
Best Local Similarity: 30.43% Mismatches: 31  
Query Match: 12.92% Indels: 35  
DB: 10 Gaps: 6

US-10-087-573-2 (1-141) x BE284887 (1-640)

QY 3 SerThrSerThrThrThrAsnPhValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
DB 444 TCCAGGTGACGACAGGCTCCCGAAGGCGACGACCCGCCATCCGGATCTCCAC 385  
QY 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
DB 384 ACCATA-----ACTGTGCAGTCTCTCAGAGCCACTGGCAATGACATTG 343  
QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
DB 342 TCATTG-----TGTGGACACAGCGGATCTAGCAGAGGCGAGTGTGCCCGCAG 292  
QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
DB 291 ACCAGGGCGACGTTCTTGTCTACTCGTCCAGTCTTGCCTAGGGTAGCACCAGGAAGGCC 232  
QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
DB 231 COTCCCCCA-----CTGGCCTCAGATC 208  
QY 95 ArgGluMetAsn-----LysThrIleSerGln 103  
DB 207 AGAGCCATGAAGTGGGGTTGACAGCGCAGAGGACCACTGTCCCAA 163

RESULT 33  
BE289370/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BE289370 650 bp mRNA linear EST 26-OCT-2000  
601089338F1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:3484185 5',  
mRNA sequence.  
BE289370 BE289370.1 GI:9169696  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 650)  
NIH-MGC <http://mgi.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-rc@mail.nih.gov](mailto:cgabs-rc@mail.nih.gov)  
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM8516 row: 1 column: 10  
High quality sequence stop: 589.  
Location/Qualifiers  
1..650  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3484185"

FEATURES  
source

/tissue\_type="tumor, gross tissue"  
 /dev\_stage="7 months"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI CGAP Mam5"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: Sali;  
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
 Library constructed by Life Technologies. Investigators  
 providing samples: Lothar Hennighausen/Robin Humphreys,  
 NIH"

BASE COUNT 107 a 197 c 210 g 136 t

## ALIGNMENT SCORES:

Pred. No.: 35.8 Length: 650  
 Score: 91.50 Matches: 35  
 Percent Similarity: 42.61% Conservative: 14  
 Best Local Similarity: 30.43% Mismatches: 31  
 Query Match: 12.92% Indels: 35  
 DB: 10 Gaps: 6

US-10-087-573-2 (1-141) x BE289370 (1-650)

QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
 |||||  
 DB 457 TCCAGGTGACGACGAGCTCCCGCAAGGCACACAGGCCCCCATCCGGATCTCCAC 398  
 |||||  
 QY 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
 |||||  
 DB 397 ACCATA-----ACTGTGAGTCTCTCAGAGCCACTGGCAATGACATTG 356  
 |||||  
 QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
 |||||  
 DB 355 TCATTG-----TGTGGACACAGGCGATGTTAGCACAGGGGCGAGTGTGGCGGAG 305  
 |||||  
 QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
 |||||  
 DB 304 ACCAGGGGACGCTCTGTCTACTCTCCAGTCTTGCCTAGGGTAGCACAGGAAGGCC 245  
 |||||  
 QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysLeu 94  
 |||||  
 DB 244 CCTCCCCA-----CTGGCCTCACAGATC 221  
 |||||

RESULT 34  
 BI150185/c  
 LOCUS 602848591F1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:5011991 5',  
 DEFINITION mRNA sequence.

ACCESSION BI150185  
 VERSION BI150185.1 GI:14610186  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)

ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 674)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LHAM1062 row: f column: 24  
 High quality sequence start: 2

High quality sequence stop: 674.

FEATURES  
 source  
 1..674  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="CZECH II"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5011991"  
 /tissue\_type="spontaneous tumor, metastatic to mammary."  
 Stem cell origin  
 /lab\_host="DH10B"  
 /clone\_lib="NCI CGAP Lu29"  
 /note="Organ: lung; Vector: pCMV-SPORT6; Site 1: Sali;  
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
 Library constructed by Life Technologies. Investigator  
 providing samples: Gilbert Smith, NIH"

BASE COUNT 120 a 192 c 224 g 138 t

## ORIGIN

Alignment Scores:  
 Pred. No.: 37.5 Length: 674  
 Score: 91.50 Matches: 35  
 Percent Similarity: 42.61% Conservative: 14  
 Best Local Similarity: 30.43% Mismatches: 31  
 Query Match: 12.92% Indels: 35  
 DB: 12 Gaps: 6

US-10-087-573-2 (1-141) x BI150185 (1-674)

QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
 |||||  
 DB 451 TCCAGGTGACGACGAGCTCCCGCAAGGCACACAGGCCCCCATCCGGATCTCCAC 392  
 |||||  
 QY 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
 |||||  
 DB 391 ACCATA-----ACTGTGAGTCTCTCAGAGCCACTGGCAATGACATTG 350  
 |||||  
 QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
 |||||  
 DB 349 TCATTG-----TGTGGACACAGGCGATGTTAGCACAGGGGCGAGTGTGGCGGAG 299  
 |||||  
 QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
 |||||  
 DB 298 ACCAGGGGACGCTCTGTCTACTCTCCAGTCTTGCCTAGGGTAGCACAGGAAGGCC 239  
 |||||  
 QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysLeu 94  
 |||||  
 DB 238 CCTCCCCA-----CTGGCCTCACAGATC 215  
 |||||

## RESULT 35

BI581991/c

LOCUS 602099446F1 NCI\_CGAP\_Co24 Mus musculus cDNA clone IMAGE:4219522 5',  
 DEFINITION mRNA sequence.

ACCESSION BI581991

VERSION BI581991.1 GI:11655612

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 727)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished

CONTACT: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

Plate: L1AM9801 row: k column: 11  
High quality sequence stop: 707.  
Location/Qualifiers

## FEATURES

source

1. 727  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4219522"  
/lab\_host="NCI CGAP Co24"  
/clone\_lib="NCI CGAP Mam2"  
/note="Organ: colon; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.6 kb. Constructed by Life  
Technologies. Note: this is a NCI CGAP Library."  
BASE COUNT 132 a 216 c 225 g 154 t  
ORIGIN

Alignment Scores: 41.3 Length: 727  
Pred. No.: 91.50 Matches: 35  
Score: 42.61% Conservative: 14  
Percent Similarity: 30.43% Mismatches: 31  
Best Local Similarity: 12.92% Indels: 35  
Query Match: 10 Gaps: 6  
DB:

US-10-087-573-2 (1-141) x BF581991 (1-727)

QY 3 SerThrSerThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22

Db 458 TCCAGGTGACGACAGGCTCCCGCAGGCGACGACCCGCCATCCGGGATCTCCAC 399

QY 23 AspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42

Db 398 ACCATA-----ACTGTGCAGTCTCCAGAGCCACTGGCAATGACATTG 357

QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58

Db 356 TCATTG-----TGTGGACACAGCGCATGCTAGCACAGGGGAGTGGCCGCGAG 306

QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74

Db 305 ACCAGGGGCGACGTTCTTGTCTACTCGTCCAGTCTTGCTAGGGGTAGCACGAGGAGGCC 246

QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94

Db 245 CTCTCCCA-----CTGGCTCACAGATC 222

QY 95 ArgGluMetAsn-----LysThrIleSerGln 103

Db 221 AGAGCCATGAACCTGGGGTTGACAGCGCAGAGCCACTGTCCCAA 177

RESULT 36

BI696433/c

LOCUS 603345762P1 NCI CGAP Mam2 Mus musculus cDNA clone IMAGE:5373471 5',

DEFINITION mRNA sequence.

ACCESSION BI696433

VERSION BI696433.1 GI:15659062

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

1 (bases 1 to 768)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

## COMMENT

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L1AM91950 row: l column: 16

High quality sequence stop: 753.

## FEATURES

source

1. 768  
Location/Qualifiers

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="FVB/N-3"

/db\_xref="taxon:10090"

/clone="IMAGE:5373471"

/tissue\_type="tumor, biopsy sample"

/dev\_stages="5 months"

/lab\_host="DH10B"

/clone\_lib="NCI CGAP Mam2"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;

Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.

Library constructed by Life Technologies. Investigator

providing samples: Gilbert Smith, NIH"

BASE COUNT 139 a 225 c 243 g 161 t

ORIGIN

Alignment Scores: 44.3 Length: 768  
Pred. No.: 91.50 Matches: 35  
Score: 42.61% Conservative: 14  
Percent Similarity: 30.43% Mismatches: 31  
Best Local Similarity: 12.92% Indels: 35  
Query Match: 12 Gaps: 6  
DB:

US-10-087-573-2 (1-141) x BI696433 (1-768)

QY 3 SerThrSerThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22

Db 455 TCCAGGTGACGACAGGCTCCCGCAGGCGACGACCCGCCATCCGGGATCTCCAC 396

QY 23 AspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42

Db 395 ACCATA-----ACTGTGCAGTCTCCAGAGCCACTGGCAATGACATTG 354

QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58

Db 353 TCATTG-----TGTGGACACAGGCGCATGCTAGCACAGGGGAGTGGCCGCGAG 303

QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74

Db 302 ACCAGGGGCGACGTTCTTGTCTACTCGTCCAGTCTTGCTAGGGGTAGCACGAGGAGGCC 243

QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94

Db 242 CTCTCCCA-----CTGGCTCACAGATC 219

RESULT 37

BI158106/c

LOCUS 602920741P1 NCI CGAP Mam3 Mus musculus cDNA clone IMAGE:5061010 5',

DEFINITION mRNA sequence.

ACCESSION BI158106

VERSION BI158106.1 GI:14618107

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 786 bp mRNA linear EST 05-JUL-2001

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 786)  
NIH-MGC <http://imgc.nhl.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov  
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM1165 row: a column: 11  
High quality sequence stop: 764.

<b>FEATURES</b>	<b>SOURCE</b>
-----------------	---------------

```

1. :786
/organism="Mus musculus"
/mol_type="mRNA"
/strain="129,CS7BL/6J,FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5061010"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/lab_host="NCI_CGAP_Mam3"
/notice="organ: mammary; Vector: pCMV-
Site 2: NCI; Cloned unidirectionally.
Library constructed by Life Technologies
providing samples: Lothar Hennighaus
Reference for transgenic model: Xu et
22, 37-43 (1999).".
146 a 230 c 245 g 165 t
cores:
45.7 Length: 786
91.50 Matches: 35
Similarity: 42.61% Conservative: 14
30.43% Mismatches: 31
12.92% Indels: 35
12 Gaps: 6

```

BASE COUNT	146 a	230 c	245 d
22, 31, 35 (100%)			

BRITAIN

Alignment Scores:		
Align. NO.:	45	7
Score:	91	50
Matches:	32	35
Percent Similarity:	42	61
Best local Similarity:	30	43
Best local Similarity:	30	43
Query Match:	12	92
DB:	12	
Gaps:	1	6
Indels:	3	5
Mismatches:	3	5
Conservative:	14	
Matches:	35	786
Length:		

US-10-092-573-2 (1-141) x BT158106 (1-786)

Qy	3	SerThrSerThrThrThrAsn	PheValAlaGluAsnArgProThrPheGlyGluThrPhe	22
Db	456	TCCAAGGTGACGACAGCGTCTCCGCAAGGGCAGACAC	CGGCCCCCATCCGGGATCTCCAC	397
Qy	23	AspValMetArgGluAlaLeu	ArgValIysSerGluArgLeuAlaMetLeuArg	42
Db	396	ACCAT	-----ACTTGCAGTCTCTCAGAGCCATCGGCATGCATTC	355
Qy	43	AlaLeuAlaGlyMetCysGlyHis	ArgValLeuProGlyThrGlyAla	58
Db	354	TCATTG	-----TGTGCACACCGGCATGTCTAGCAGGGCAGGTGGCCGAC	304
Qy	59	-----SerAlaIleAlaIa	Thr-----ValThrProLysGlyAlaSerMetLysLeu	74
Db	303	ACCAGGGCAGCTTCTGTCTACTCGTCCAGTCTTGC	TAGGGGTAGCACCGAAGGCC	244
Qy	75	LysProArgProGlnSerThrLys	SerProGluLeuArgGluLeuSerArgLysIle	94
Db	243	CCTCCCCA	-----CTGGCTTCACATC	220
Qy	95	ArgGluMetAsn	-----LysThrIleSerGln	103
Db	219	AGACCCATGA	CTTGGGTTGCACAGCGAGAAGCCACTGTCCCAA	175

RESULT 38  
BI839165/C

LOCUS	BI839165	835 bp	mRNA	linear	EST 04-OCT-2001
DEFINITION	603086487F1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5225661 5', mRNA sequence.				
ACCESSION	BI839165				
VERSION	BI839165.1 GI:15950715				
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 835)				
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished				
COMMENT	Contact: Robert Strausberg, Ph.D.				

Email: [cs9abbs@mail.nih.gov](mailto:cs9abbs@mail.nih.gov)  
 Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM11567 row: m column: 22  
 High quality sequence start: 22  
 High quality sequence stop: 537.

FEATURES  
source

```

1. .e33
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5225661"
/lab_host="DH10B"
/clone_lib="NIH_MGC_120"
/notes="Organ: pooled pancreas and spleen; Vector:
pCMV-Sport6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA
source anonymous pool of spleen and pancreas from 28 YO
male. Library is oligo-dT primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 1.5 kb, insert size range 1-2.5 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 035. Note: this is a NIH MGC Library."

```

BASE COUNT	160 a	245 c	254 q	176 t
tracking code 025. Note: CH				

ORIGIN

```

ORIGIN
Alignment Scores:
Pred. No.: 49.3 Length: 835
Score: 91.50 Matches: 42
Percent Similarity: 38.41% Conservative: 16
Best Local Similarity: 27.81% Mismatches: 34
Query Match: 12.9% Indels: 59
DB: 12 Gaps: 9

US-10-087-573-2 (1-141) x BI839165 (1-835)
Qy 15 ArgProThrPheGlyGluThrPheAspVal-----MetArgGluAlaLeu--- 29
Db 538 AGGCCCACCATGGCCACACGCTTGGTGTGACCCCTCCAGGGGTGACAGCGGGCTCCG 479
Qy 30 -----LeuArgValIysSerSer 35
Db 478 CAGGGGCAGCATCAGGCCCCCATCCCGGGATCTCCACACCATGACTGTGCAGCTCTG 419
Qy 36 GluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyVHisArgValLeuProGly 55
Db 418 GAGCCACTGGCAATGACGTTGTCATTG-----TCCGGGCACCCAGCGCATGTCTAGC 368
Qy 56 ThrGlyAla-----SerAlaLeuAlaAlaThr-----ValThrPro 67
Db 367 ACAGGGGCTGTGTGGCCACACAGCGTGGGCGCATTTCTTGTCCACAGTCCAGTCTTGGCC 308
Qy 68 IysGValAlaSerMetIysLeuLysProProArgProGlnSerThrLysSerProGluLeu 87

```

```
Db 307 AGGGGACGACACGAGAGGCCCTCCCGG----- 278
Qy 88 ArgGluLeuSerArgLysIleArgGluMetAsn-----LysThrIle 101
Db 277 -----CTGGCTCACAGATCAGGCGCCAACTTAGGGTTGACAGCACAGAGCCACTG 224
Qy 102 SerGlnGluSerAlaArgValAsnHis-----ArgLeu 112
Db 223 TCCCAAGTGGTCTGTGAGACAGCGCACATCTTCATAGCACTGGTGGCGGCTGGCGGCTG 164
Qy 113 ProGlu-----GlyHisProLeuLeuLys 121
Db 163 TCCCAACACGTCGGCGGAACCTGCTGGAGCGG 131

RESULT 39
BI411683/c 849 bp mRNA linear EST 14-AUG-2001
LOCUS 602965493F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5120758 5',
DEFINITION mRNA sequence.
ACCESSION BI411683
VERSION BI411683.1 GI:15172606
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 849)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rc@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1294 row: j column: 23
High quality sequence start: 3
High quality sequence stop: 794.
Location/Qualifiers
1..849
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5120758"
/tissue_type="pooled lung tumors"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NCI_CGAP_Lu33"
/notes="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site 1: NotI; Site 2: EcoRI; 1st
strand cDNA was prepared from mRNA obtained from pooled
lung tumors with a Not I - oligo(dT) primer [5',
TGTATCCATCTGAGTGGAGCGCGCCCTGTTTCTTTTCTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
```

```
BASE COUNT 163 a 232 c 272 g 182 t
ORIGIN
Alignment Scores:
Pred. No.: 50.4 Length: 849
Score: 91.50 Matches: 35
Percent Similarity: 42.61% Conservative: 14
Best Local Similarity: 30.43% Mismatches: 31
Query Match: 12.92% Indels: 35
```

```
Db 12 Gaps: 6
US-10-087-573-2 (1-141) x BI411683 (1-849)
Qy 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22
Db 351 TCCAAGTGGAGGACGCTCCCGCAAGGCGACGACAGGCCCCCCTCCGGATCTCCAC 292
Qy 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42
Db 291 ACCATA-----ACTGTGCAGTCTCTCAGAGCCACTGGCAATGACATTG 250
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58
Db 249 TCATTG-----TGTGGACACAGCGCGATCTAGCACAGGCGCAGTGTGGCGCGAG 199
Qy 59 -----SerAlaIleAlaIleThr-----ValThrProLysGlyAlaSerMetLysLeu 74
Db 198 ACCAGGGCGACGTTCTTGTCTACTCTCCAGTCTTCCTAGGGGTAGCAGGAGAGGCC 139
Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94
Db 138 CTCCCCCA-----CTGGCTCACAGATC 115
Qy 95 ArgGluMetAsn-----LysThrIleSerGln 103
Db 114 AGAGCCATGAACCTGGGGTTGACAGCGCGACAGCCACTGTCCCAA 70

RESULT 40
CB586950/c 872 bp mRNA linear EST 03-APR-2003
LOCUS AGENCOURT_12989871 NIH_MGC_136 Mus musculus cDNA clone
DEFINITION IMAGE:30292954 5', mRNA sequence.
ACCESSION CB586950
VERSION CB586950.1 GI:29504806
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 872)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rc@mail.nih.gov
Tissue Procurement: Dr. David Rowe
cDNA Library Preparation: Invitrogen Corp
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM337 row: k column: 11
High quality sequence stop: 661.
Location/Qualifiers
1..872
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30292954"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_136"
/notes="Vector: pCMV-SPORT6.1 ccdB; Site 1: EcoRV; Site 2:
NotI; Normalized, full-length enriched library from pool
of mouse embryonic limb, maxilla and mandible, embryonic
day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and
jaw equivalents from respective days). Cloned
directionally, oligo-dT primed
(5'-GACTAGTCTTCTAGATCGGAGCGCGCCCTT)15-3'. Size selected
for the 21kb fragments, average insert size 1.2 kb.
Normalization to Cot 7.5. Tissue contributed by David
```

Rowe; library constructed by ResGen, Invitrogen Corp.  
Note: this is a NIH/MGC Library."

BASE COUNT 165 a 253 c 274 g 180 t  
ORIGIN

Alignment Scores:  
Pred. No.: 52.1 Length: 872  
Score: 91.50 Matches: 35  
Percent Similarity: 42.61% Conservative: 14  
Best Local Similarity: 30.43% Mismatches: 31  
Query Match: 12.92% Indels: 35  
DB: 14 Gaps: 6

US-10-087-573-2 (1-141) X CB586950 (1-872)

Qy 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
Db 443 TCCAAGTGACGACAGGCTCCCGAAGGGCAGCACCCAGGCCCCCATCCGGGATCTCCAC 384  
Qy 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgGluAlaMetLeuArg 42  
Db 383 ACCATA-----ACTGTGAGTCTCTCAGAGCCACTGGCAATGACATTG 342  
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
Db 341 TCATTG-----TGTGGACACCGGCGATGCTAGCACAGGGCGCAGTGTGGCGCGAG 291  
Qy 59 -----SerAlaIleAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
Db 290 ACCAGGGGCACGTTCTGTCTACTCTGTCAGTCTTCCCTAGGGGTAGCACAGGAAGGCC 231  
Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
Db 230 CTTCCCTCCA-----CTGGCCTCACAGATC 207  
Qy 95 ArgGluMetAsn-----LysThrIleSerGln 103  
Db 206 AGAGCCATGAACCTGGGGTTGACAGCGCAGAGCCACTGTCCCAA 162

Search completed: November 17, 2003, 14:11:40  
Job time : 1938 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 10:38:30 ; Search time 41 Seconds  
(without alignments)

545.865 Million cell updates/sec

Title: US-10-087-573-2

Perfect score: 708

Sequence: 1 MESTSTTNFAVNRPTGE.....RAEYRHLRLSKQGVNELI 141

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

1:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	708	100.0	141	23 ABP53714	Babesia canis Bcvi
2	638	90.1	285	23 ABP53715	Babesia canis Bcvi
3	83	11.7	192	7 AAP63504	Sequence of p' pro
4	81.5	11.5	945	23 AAU03538	Human protein kina
5	81.5	11.5	945	23 AAU03537	Human kinase polyp
6	81.5	11.5	1259	23 ABG61917	Prostate cancer-as
7	80.5	11.4	1157	22 ABB63351	Drosophila melanog
8	79	11.2	192	20 AAY01973	HIV-1 viral infect
9	79	11.2	1780	22 ABUS3201	Human cell cycle-a

10	79	11.2	1780	22	AA338681	Human polypeptide
11	79	11.2	1788	22	AA40467	Human polypeptide
12	78.5	11.1	828	22	ABG18310	Novel human diago
13	78.5	11.1	828	22	ABG29103	Novel human diago
14	78	11.0	203	7	AAP63348	HTLV-III virus (H
15	77.5	10.9	409	22	AB88134	Drosophila melanog
16	77.5	10.9	451	23	AB883793	Fungal decaprenyl
17	77	10.9	96	20	AAW94135	VIF-derived HIV pr
18	77	10.9	192	20	AAW01976	HIV-1 viral infect
19	77	10.9	192	20	AAW01977	HIV-1 viral infect
20	77	10.9	192	20	AAW01978	HIV-1 viral infect
21	77	10.9	192	20	AAW01970	HIV-1 viral infect
22	77	10.9	192	20	AAW01971	HIV-1 viral infect
23	77	10.9	192	20	AAW01972	HIV-1 viral infect
24	77	10.9	192	20	AAW89324	HIV-1 P protein se
25	77	10.9	192	22	AAW85994	Amino acid sequenc
26	77	10.9	203	7	AAP60421	Sequence of LAV vi
27	77	10.9	203	14	AAW43868	HTLV-III SOR gene
28	77	10.9	203	23	AAO19388	Lymphadenopathy-as
29	76.5	10.8	214	22	AAW3854	Human prostate can
30	76.5	10.8	373	20	AAW50122	Human histone fusi
31	76.5	10.8	373	22	AAW30626	Amino acid sequenc
32	76.5	10.8	1778	22	AAW79480	Human protein SEQ
33	76.5	10.8	2099	22	AAW78496	Human protein SEQ
34	76	10.7	203	14	AAW43876	HTLV-III SOR gene
35	76	10.7	259	22	AAW43752	Human prostate can
36	75.5	10.7	442	22	AB88120	Drosophila melanog
37	75.5	10.7	542	22	ABG19829	Novel human diago
38	75.5	10.7	1257	24	ABJ19818	Androgen-independe
39	75	10.6	476	22	ABG62927	Drosophila melanog
40	74.5	10.5	1116	22	AB88597	Drosophila melanog
41	74	10.5	192	20	AAW90177	HTLV-III p' protei
42	74	10.5	502	23	AB883782	Fungal decaprenyl
43	73.5	10.4	190	20	AAW01980	Consensus sequence
44	73.5	10.4	191	20	AAW01980	HIV-1 viral infect
45	73.5	10.4	192	20	AAW01979	HIV-1 viral infect

#### ALIGNMENTS

RESULT 1  
ABP53714  
ID ABP53714 standard; Protein; 141 AA.  
AC ABP53714;  
DT 23-DEC-2002 (first entry)  
DE Babesia canis Bcvi15 15kd protein SEQ ID NO:2.  
KW Babesia canis; Bcvi15; 15kd protein; Bcvi32; 32kd protein; infection;  
XX antiparasitic; immunostimulant; vaccine.  
XX Babesia canis.  
XX EPI238983-A1.  
XX 11-SEP-2002.  
XX 04-MAR-2002; 2002EP-0075930.  
XX 06-MAR-2001; 2001EP-0200816.  
XX (ALKU ) AKZO NOBEL NV.  
XX Schettlers TPM, Carcy BPD, Drakulovski PR, Gorenflot AF;  
XX WPI; 2002-724917/79.  
XX N-PSDB; ABQ82649.

Novel Babesia canis associated protein and nucleic acid encoding the protein, useful in a vaccine and in the manufacture of vaccines for

PT combating Babesia canis infections  
 XX Claim 1; Fig 2; 41pp; English.  
 PS  
 CC The present invention describes a Babesia canis associated protein (I),  
 CC comprising a BcVir15 protein of 15 kD molecular weight (MW) and having  
 CC a sequence of at least 80% homology to a sequence (ABP53715) of 285 amino acids, or  
 CC a BcVir15 protein of 32 kD MW and having a sequence of 141  
 CC amino acids, or a BcVir15 protein of 32 kD MW and having a sequence of  
 CC at least 80% homology to a sequence (ABP53715) of 285 amino acids, or  
 CC their immunogenic fragments. (I) have antiparasitic and immunostimulant  
 CC activities, and can be used in vaccines. (I) can also be used for the  
 CC preparation of a vaccine for combating B. canis infections. (I) is also  
 CC useful in a diagnostic test for the detection of antibodies against  
 CC B. canis associated antigenic material. The present sequence represents  
 CC BcVir15 from the present invention.  
 XX  
 SQ Sequence 141 AA;  
 Query Match 100.0%; Score 708; DB 23; Length 141;  
 Best Local Similarity 100.0%; Pred. No. 1,1e-68;  
 Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MESTSTTNFVAENRPTFGTETDVMREALLRVKSSERLALAGMCGHRVLPQTGASA 60  
 DB 1 MESTSTTNFVAENRPTFGTETDVMREALLRVKSSERLALAGMCGHRVLPQTGASA 60  
 QY 61 IAATVTPKGASMKLKPDPQSTKSPPELRELSRKIRENMKTIQSASRVNHRLEPGHPLE 120  
 DB 61 IAATVTPKGASMKLKPDPQSTKSPPELRELSRKIRENMKTIQSASRVNHRLEPGHPLE 120  
 QY 121 KRAEYFRHLRLSKSQ 135  
 DB 121 KRAEYFRHLRLSKSQ 135  
 RESULT 2  
 ABP53715  
 ID ABP53715 standard; Protein; 285 AA.  
 XX  
 AC ABP53715;  
 DT 23-DEC-2002 (first entry)  
 DE  
 KW Babesia canis BcVir15 32kD protein SEQ ID NO:4.  
 XX  
 KW Babesia canis; BcVir15; 15kD protein; BcVir15; 32kD protein; infection;  
 KW antiparasitic; immunostimulant; vaccine.  
 XX  
 OS Babesia canis.  
 PN EP1238983-A1.  
 XX  
 PD 11-SEP-2002.  
 XX  
 PF 04-MAR-2002; 2002EP-0075830.  
 XX  
 PR 06-MAR-2001; 2001EP-0200816.  
 XX  
 PA (ALKU ) AKZO NOBEL NV.  
 PI Schettlers TPM, Carcy BPD, Drakulovski PR, Gorenflot AF;  
 XX  
 DR WPI; 2002-724917/79.  
 DR N-PSDB; ABQ82650.  
 XX  
 PT Novel Babesia canis associated protein and nucleic acid encoding the  
 PT protein, useful in a vaccine and in the manufacture of vaccines for  
 PT combating Babesia canis infections -  
 XX  
 PS Claim 2; Fig 3; 41pp; English.  
 XX  
 CC The present invention describes a Babesia canis associated protein (I),  
 CC comprising a BcVir15 protein of 15 kD molecular weight (MW) and having

CC a sequence of at least 80% homology to a sequence (ABP53714) of 141  
 CC amino acids, or a BcVir15 protein of 32 kD MW and having a sequence of  
 CC at least 80% homology to a sequence (ABP53715) of 285 amino acids, or  
 CC their immunogenic fragments. (I) have antiparasitic and immunostimulant  
 CC activities, and can be used in vaccines. (I) can also be used for the  
 CC preparation of a vaccine for combating B. canis infections. (I) is also  
 CC useful in a diagnostic test for the detection of antibodies against  
 CC B. canis associated antigenic material. The present sequence represents  
 CC BcVir15 from the present invention.  
 XX  
 SQ Sequence 285 AA;  
 Query Match 90.1%; Score 638; DB 23; Length 285;  
 Best Local Similarity 94.8%; Pred. No. 1e-60;  
 Matches 128; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 MESTSTTNFVAENRPTFGTETDVMREALLRVKSSERLALAGMCGHRVLPQTGASA 60  
 DB 1 MESTSTTNFVAENRPTFGTETDVMREALLRVKSSERLALAGMCGHRVLPQTGASA 60  
 QY 61 IAATVTPKGASMKLKPDPQSTKSPPELRELSRKIRENMKTIQSASRVNHRLEPGHPLE 120  
 DB 61 IAATVTPKGASMKLKPDPQSTKSPPELRELSRKIRENMKTIQSASRVNHRLEPGHPLE 120  
 QY 121 KRAEYFRHLRLSKSQ 135  
 DB 121 KRAEYFRHLRLSKSQ 135  
 RESULT 3  
 AAP61504  
 ID AAP61504 standard; Protein; 192 AA.  
 XX  
 AC AAP61504;  
 DT 08-JUN-1991 (first entry)  
 DE  
 KW HIV; LAV; AIDS; diagnosis; vaccine.  
 XX  
 OS HTLV-IIIb/H9 cells (ATCC CRL 8543).  
 XX  
 PN EPI87041-A.  
 XX  
 PD 09-JUL-1986.  
 XX  
 PF 23-DEC-1985; 85EP-0309454.  
 XX  
 PR 24-DEC-1984; 84US-0685272.  
 PR 04-DEC-1985; 85US-0805069.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Capon DJ, Lasky LA;  
 XX  
 DR WPI; 1986-177602/28.  
 DR N-PSDB; AAN60288.  
 XX  
 PT Acquired immune deficiency syndrome polypeptide(s) - obt'd. by  
 PT molecular cloning etc. and used for diagnosis and in vaccines  
 PT against virus disease  
 XX  
 PS Example; fig 2; 125pp; English.  
 XX  
 CC A comparison of AAN60287 with the cDNA of the HTLV-III genome  
 CC revealed one particular clone, designated p7.11 which contained a  
 CC DNA sequence encoding this peptide (AAP60308) sequence. This approx.  
 CC 2.2 kilobase covers the precursor gag region and encodes, 5' to 3',  
 CC p-12, p-15, p-24 a second p-15 protein and approx. 300 extra base  
 CC pairs 3' to the gag region (see AAN60288).  
 XX  
 SQ Sequence 192 AA;